THE JOURNAL OF
EDUCATIONAL PSYCHOLOGY

Volume XXVIII December, 1937 Number 9

THE VALIDITY OF CERTAIN ADJUSTMENT INVENTORIES IN PREDICTING PROBLEM BEHAVIOR

NOEL KEYS

University of California, Berkeley, California

AND

MARGARET S. GUILFORD

Orland High School, Orland, California

Broadening concepts of education increasingly stress the function of the school in individual guidance and adjustment. This in turn has quickened the demand for devices to aid in identifying pupils in need of special attention, and analyzing the difficulties which confront them. Many tests, inventories, rating scales, and record systems are now available for this purpose. The better of these are the fruit of extensive experimentation and validation by their respective authors. Only rarely, however, have the finished instruments been submitted to objective investigation to determine their comparative merits for particular uses.

The present study undertakes to compare the serviceability of four inventories of the questionnaire type in predicting problem behavior at ninth- and tenth-grade levels. The behavior in question may be defined as "disciplinary conduct resulting from social and emotional maladjustment." Important contributory factors would probably be found in intellectual deficiencies, poor home environment, ill health, and emotional instability. With the exception of intelligence, the relation of which to misbehavior has already been studied extensively, each of the factors named is measured by one or more of the testing instruments considered.

TESTS OR INVENTORIES COMPARED

The four inventories selected for study are commercially available and favorably known. Since the facts as to their derivation and

standardization have been adequately set forth by their authors in the test manuals or supplementary articles, brief descriptions may here suffice:

1. The Sims Score Card for Socio-Economic Status[1] consists of a folder of twenty-three questions covering such matters as the number of rooms and conveniences in the home, parents' education and occupation, family cultural interests, recreations, etc. Pupils' answers yield scores which are reported to show no overlapping between children of professional parents and those of unskilled laborers.[2] Obviously such scores are not a direct measure of behavior or adjustment. Delinquency is, however, frequently attributed to the living conditions of children in "underprivileged homes." Moreover, a questionnaire similar to this was one of the only two devices found by Lentz[3] to discriminate repeatedly between delinquent boys and an unselected public-school group.

2. The Bell Adjustment Inventory[4] is composed of one hundred ten questions to be answered "Yes," "No," or "?." These yield separate scores for home, health, social, and emotional adjustments. The following are typical of questions in the four fields: (Home) "Does either of your parents become angry easily?" (Health) "Have you frequently been absent from school because of illness?" (Social) "Do you ever cross the street to avoid meeting somebody?" (Emotional) "Does it frighten you to be alone in the dark?" High scores indicate unsatisfactory adjustments. Individuals scoring high on social items, for example, are said by the manual to tend to be submissive and retiring; and those scoring high on emotional items to be unstable.

3. The Personality Inventory by R. G. Bernreuter[5] includes one hundred twenty-five questions similar in form to those employed by Bell. By means of separate scoring keys assigning different weights to answers, the same paper can be made to yield four distinct scores,

---

[1] Sims, V. M.: *Sims Score Card for Socio-Economic Status, Form C.* Bloomington, Ill.: Public School Publishing Co., 1927.

[2] Sims, V. M.: *The Measurement of Socio-Economic Status.* Public School Publishing Co., 1928, p. 29.

[3] Lentz, T. F., Jr.: *An Experimental Method for the Discovery and Development of Tests of Character.* New York: Teachers College, Columbia University, Contributions to Education, No. 180, 1925, *ad loc.*

[4] Bell, H. M.: *The Adjustment Inventory.* Stanford University, Calif.: Stanford University Press, 1934.

[5] Bernreuter, R. G.: *The Personality Inventory.* Stanford University Press, 1931.

termed B1-*N* for neurotic tendency, B2-*S* for self-sufficiency, B3-*I* for introversion, and B4-*D* for dominance. High scores are said to indicate, respectively, neurotic tendency or emotional instability, a preference for being alone and avoiding sympathy or advice, introversion, and a tendency "to dominate others in face-to-face situations." Since the scale for introversion is reported in the manual to correlate .94 with that for neuroticism, which has a higher reliability, the papers of the present study were not scored for introversion.

4. The Personal Index by G. C. Loofbourow and Noel Keys[1] is designed, according to the manual, "for the detection of attitudes indicative of problem-behavior." As published in 1933, this had four parts: a "vocabulary" test containing thirty fictitious words concerning which the subject is to state whether he knows their meaning, a multiple-choice test in which he is to check the answer best expressing his attitude toward "bumming rides," "speed-cops," and fifteen other practices or persons, a "virtues" test of twenty-one questions, and an adjustment questionnaire of eighty-nine items similar to those of the Bell and Bernreuter inventories. Later, unpublished studies by A. E. Joyal and W. C. Riggs in Denver, and Keys in Pacific Coast cities indicated that the "virtues" test tended to lower the validity of total scores. This section was consequently omitted from the 1937 edition of the Index. In the present study, scores were computed separately according to both 1933 and 1937 forms.

All four of the instruments named are reported by their authors to show reliabilities above .90, and reasonably satisfactory validity, according to various criteria.

SUBJECTS OF THE EXPERIMENT

The freshman and sophomore classes of the high school of Chico, California, constituted the subjects of the present experiment. This is a four-year school of near one thousand enrolment, located in the upper Sacramento Valley, some one hundred ninety miles from San Francisco. The study was made possible only by the prolonged coöperation of Principal F. L. Cummings and his staff of counselors and homeroom teachers. With their assistance, the tests named were administered to the two classes in the school year 1934–1935. The sole exception was the Bernreuter Inventory, which had already been given to members of the sophomore class in their freshman year. The

[1] Loofbourow, G. C., and Keys, Noel: *Personal Index.* Minneapolis, Minn.: Educational Test Bureau, 1933. Revised edition, 1937.

Bernreuter scores of these pupils were, therefore, copied from the school records.

Complete data on all four tests were secured for one hundred forty boys and one hundred thirty girls. An additional fifty-seven boys and seventy-eight girls, however, had records complete save for one test, missed by reason of absence or the like. It was decided to include these cases also, making a total of one hundred ninety-seven boys and two hundred eight girls, or four hundred five pupils in all. Owing to the fact that in a few homerooms satisfactory teacher ratings proved impossible to obtain, numbers in computations involving the Haggerty-Olson-Wickman scale were still further reduced. As shown in the ensuing tables, correlations were based on numbers ranging from one hundred thirty to three hundred seventy-five. As between one inventory and another, however, differences in the populations involved were small, and probably did not appreciably affect the results.

CRITERIA OF BEHAVIOR

In evaluating pupils' behavior, two distinct types of rating were employed. Of these, one involved highly detailed and itemized ratings by single judges; the other, a blanket rating determined by the composite judgment of a group. For the first, the Haggerty-Olson-Wickman Behavior Rating Schedule[1] was utilized. This device entails five pages of ratings for each individual, with subsequent computation of scores. Schedule *A* is concerned with the comparative frequency of fifteen types of overt misconduct, ranging from tardiness and cheating to stealing and sex offenses. Schedule *B* involves placing the pupil by check mark on graphic and well-defined scales for thirty-five traits such as alertness, courtesy, energy, excitability, and moodiness. Scores are the sum of weights experimentally determined for behavior of the type noted with respect to each of the fifty points considered.

To determine whether ratings on overt misbehavior of the type covered by Schedule *A* would constitute a more satisfactory criterion than the totals for *A* and *B* combined, the writers computed ten correlations between Schedule *A* and scores on the inventories of this study. These correlations, however, averaged but .14 as compared with .18 for the corresponding correlations based on *A* and *B* totals. The total ratings were therefore adopted as the more indicative.

[1] Haggerty, M. E., Olson, W. V., and Wickman, E. K.: *Behavior Rating Schedules*. Yonkers on Hudson, N. Y.: World Book Co., 1930.

In Chico High School, entering freshmen are divided according to sex, then subdivided into groups of twenty to thirty pupils. Each such group is assigned to a particular homeroom teacher, who continues in that relation to them until their graduation. In the present study, every homeroom teacher of a freshman or sophomore group was asked by the principal to fill out Haggerty-Olson-Wickman schedules for each boy or girl in his charge. Such ratings were, in fact, obtained from eighteen teachers, ten of boys', and eight of girls' groups. These ratings were made between June 1 and December 20, 1935, or on an average of six to nine months after the inventories and tests had been administered. The time element was, therefore, such that each rater had acted as homeroom teacher to the pupils rated for a minimum of one school year; while in half the cases, the period amounted to four semesters or more.

At best, however, a rating by one individual is subject to many weaknesses. In the manual for the Haggerty-Olson-Wickman Schedules, the authors report a reliability of .92 for individual ratings as determined by split-halves and the Spearman-Brown formula, but as between the ratings of different teachers, a correlation of only "about .60." This great discrepancy clearly reflects the importance of "halo" effect and bias in individual judges. When, in addition, ratings obtained on certain pupils from one teacher are thrown into the same distribution with ratings made on other pupils by different teachers, the tendency of some teachers to rate *generally* high and others *generally* low operates to minimize the significance of the final rank order. The reliability of the resulting melange is, therefore, still further reduced. Since this was the situation in the present study, it was to be expected that the reliability of the H-O-W ratings would be under, rather than over .50. Yet the usual expedient of obtaining several ratings on each individual and averaging was here impractical, since each had had but one homeroom teacher in high school.

Fortunately, the principal, counselors, and deans have sufficient contacts with pupils to acquaint them with outstanding merits and shortcomings. In order to tap the combined knowledge of these administrators, a second, or blanket-rating, scheme was devised for use by the conference method. During the week following the close of school in June, 1935, the principal, two counselors, one dean of boys and two deans of girls met and assigned to each freshman or sophomore pupil a "star rating" on behavior. In arriving at this, they worked with class rolls before them and according to the following directions:

1. Select, to the extent of fifteen to twenty per cent of the total enrolment, those pupils you agree constitute, in school or out, behavior problems of the type defined above. Star their names in red.

2. Select those whom you regard as uncommonly well adjusted and fine persons, likewise to the extent of fifteen to twenty per cent of the enrolment. Star their names in blue.

3. Out of group 1, select those whose behavior is most serious, such as a juvenile court record or conduct justifying such, near nervous breakdowns, and the like. Limit this selection to about five per cent of the total enrolment. Before their names place a second red star.

4. Out of group 2, select those whom you regard as the outstandingly finest boys or girls in school, those of best conduct, "just what you would wish a son or daughter of yours to be," etc. Before their names place a second blue star. Mark about five per cent of the total enrolment thus.

5. Then go over the sixty or seventy per cent left unstarred and agree on those you can say are above average in their adjustment and conduct, even though not strikingly superior. Mark these with a blue plus sign. Similarly, those whose behavior is below average, even though not serious problems, mark with a red minus sign.

6. The remaining twenty to forty per cent leave unmarked, as average.

7. Convert the symbols into numerical values as follows:

Behavior	Mark	Star rating	Approximate per cent
Outstandingly finest	two blue stars	1	about 5
Very superior	one blue star	2	10 to 15
Better than average	blue plus sign	3	15 to 20
Average	unmarked	4	20 to 40
Poorer than average	red minus sign	5	15 to 20
Problem behavior case	one red star	6	10 to 15
Problems of gravest character	two red stars	7	about 5

METHODS OF DETERMINING CORRELATION

The best measure of linear agreement between two ordered series is the Pearson product-moment coefficient, r. Such coefficients were computed between scores on each of the tests or inventories and both Haggerty-Olson-Wickman and star ratings. Product-moment correlation, however, does not work to optimum advantage with a scale of

but seven intervals, such as the star ratings, or in cases in which the true relationship is curvilinear. Moreover, if the object is to determine the ability to discriminate problem-behavior cases from others, bi-serial correlation might be more appropriate. This is a method of measuring agreement between two traits, one of which is an attribute with only two categories, *e.g.*, problem and non-problem, and the other a variable with numerous gradations, such as the test or inventory scores. As a check, therefore, on the Pearson coefficients and to facilitate comparisons with studies based on dichotomous classifications of delinquent *versus* non-delinquent, bi-serial correlations were also computed.

For this purpose, pupils with star ratings of five to seven were defined as problem cases, and those rated one to four as non-problems. In order to give more weight to those whose problem status was most clearly recognized, individuals rated six were tallied twice and those rated seven tallied as though they were three persons. Similarly, among the normally behaved, pupils rated four, three, two, and one were weighted one, two, three, and four times, respectively. To obtain comparable coefficients for the H-O-W ratings, pupils lying between $-.5$ and $+.5\sigma$ were tallied once, those between $-.5$ and -1.5σ, twice, those between -1.5 and -2.5σ, three times, and those below -2.5σ, four times. Likewise, pupils rating between $+1.5$ and $+2.5\sigma$ were counted as two, and those above 2.5σ, as three problem cases. Following this weighted tallying, bi-serial r's were computed by the usual formula,[1]

$$r_{bis} = \frac{M_p - M_q}{\sigma_{total}} \cdot \frac{pq}{z},$$

in which p = the fraction of the total number rated behavior problem cases

q = the fraction of the total rated non-problem, or control cases

z = the height of the ordinate at the point of division, as read from the Kelley-Wood table of the normal-probability integral.

In interpreting the coefficients of Tables I to V, it is important to remember that high numerical ratings by either H-O-W or star-rating scales indicate problem behavior. Positive correlations, there-

[1] Kelley, T. L.: *Statistical Method.* New York: Macmillan, 1923, p. 248.

fore, signify that the attribute measured by the test or inventory tends to go with undesirable rather than desirable conduct. Since in most cases, these attributes are themselves forms of maladjustment, positive correlations would be anticipated. The Sims Score Card, however, constitutes a notable exception. High scores on it mark superior socio-economic status, which might be expected to correlate negatively with misbehavior. Positive correlations, such as occur in the case of the girls, would imply that problem-behavior is more frequent in children from homes of good rather than poor economic circumstances.

CORRELATION BETWEEN INVENTORIES AND RATINGS

Tables I and II present, respectively, the product-moment and bi-serial correlations between the several inventories and the Haggerty-Olson-Wickman ratings by homeroom teachers. Tables III and IV show the corresponding correlations with star ratings assigned by the group of administrators and counselors.

TABLE I.—PRODUCT MOMENT CORRELATIONS BETWEEN INVENTORY SCORES AND HAGGERTY-OLSON-WICKMAN BEHAVIOR RATINGS BY HOMEROOM TEACHERS

Inventory	Boys		Girls		Total	
	N	*r*	*N*	*r*	*N*	*r*
Sims Score Card for Socio-Economic Status	132	−.09 ±.05	132	.08 ±.06	264	.000±.04
Bell Adjustment Inventory						
a. Home	134	.03 ±.06	149	*.24 ±.05*	283	.15 ±.04
b. Health	134	.001±.06	149	.12 ±.06	283	.06 ±.04
c. Social	134	−.15 ±.06	149	−.06 ±.06	283	−.10 ±.04
d. Emotional	134	−.03 ±.06	149	.05 ±.06	283	.03 ±.04
Total score	134	−.08 ±.06	149	.03 ±.06	283	.04 ±.04
Bernreuter Personality Intory						
N. Neurotic tendency	130	.01 ±.06	131	−.05 ±.06	261	−.04 ±.04
S. Self-sufficiency	130	.04 ±.06	131	.003±.06	261	.07 ±.04
D. Dominance	130	.08 ±.06	131	.17 ±.06	261	.15 ±.04
Loofbourow-Keys Personal Index						
Questionnaire only	133	*.28 ±.05*	151	*.22 ±.05*	284	*.23 ±.04*
Total by 1933 edition	133	*.23 ±.05*	151	*.25 ±.05*	284	*.24 ±.04*
Total by 1937 edition	133	*.28 ±.05*	151	*.26 ±.06*	284	*.27 ±.04*

TABLE II.—MODIFIED BI-SERIAL CORRELATIONS BETWEEN INVENTORY SCORES AND HAGGERTY-OLSON-WICKMAN BEHAVIOR RATINGS BY HOMEROOM TEACHERS

Inventory	Boys		Girls		Total	
	N	r_{bis}	*N*	r_{bis}	*N*	r_{bis}
Sims Score Card for Socio-Economic Status	132	$-.12 \pm .06$	132	$.21 \pm .07$	264	$.04 \pm .05$
Bell Adjustment Inventory						
a. Home	134	$.03 \pm .07$	149	$.28 \pm .06$	283	$.17 \pm .04$
b. Health	134	$-.02 \pm .07$	149	$.11 \pm .06$	283	$.04 \pm .05$
c. Social	134	$-.24 \pm .07$	149	$-.03 \pm .06$	283	$-.04 \pm .05$
d. Emotional	134	$-.06 \pm .07$	149	$.05 \pm .06$	283	$-.003 \pm .05$
Total score	134	$-.14 \pm .07$	149	$.13 \pm .06$	283	$.01 \pm .05$
Bernreuter Personality Inventory						
N. Neurotic tendency	130	$-.09 \pm .07$	131	$-.06 \pm .07$	261	$-.12 \pm .05$
S. Self-sufficiency	130	$-.08 \pm .07$	131	$.21 \pm .07$	261	$.08 \pm .05$
D. Dominance	130	$.22 \pm .06$	131	$.32 \pm .06$	261	$.27 \pm .04$
Loofbourow-Keys Personal Index						
Questionnaire only	133	$.40 \pm .06$	151	$.30 \pm .06$	284	$.34 \pm .04$
Total by 1933 edition	133	$.36 \pm .06$	151	$.35 \pm .06$	284	$.35 \pm .04$
Total by 1937 edition	133	$.37 \pm .06$	151	$.34 \pm .06$	284	$.35 \pm .04$

In general, ratings obtained from the Haggerty-Olson-Wickman schedules (Tables I and II) correlate less well with the inventories than ratings by means of the star scale (Tables III and IV). The average of the twenty-four *r*'s for the total group, disregarding signs, is only .13 for the H-O-W, as against .21 for the star ratings. This is doubtless an indication of higher reliability and validity for the star ratings. No conclusion can be drawn, however, as to relative merits of the two devices, or of blanket *versus* itemized ratings in general, since the star ratings represent the composite judgment of a group.

Comparison of Tables I and III with II and IV, respectively, reveals that the modified bi-serial technique yields somewhat higher coefficients than does the product-moment method with these data. For H-O-W ratings, the arithmetic average of the *r*'s for the total group is .15 by the bi-serial as compared with .12 by the product-moment method, an increase of .03. As applied to the star-ratings, however, the corresponding averages are .16 and .27, a difference of .11. The much greater increase observable in the case of star-ratings suggests that product-moment correlation does not fully reflect the degree of

association when data are based on a scale of but seven intervals. A slight curvilinearity in the relation of star ratings to certain inventories was likewise overcome by this means.

The most striking features of all four tables is the generally low value of the *r*'s obtained. A correlation amounting to four times its probable error would not occur one time in one hundred by chances of sampling alone, and so may be regarded as statistically significant. Coefficients which meet this requirement are shown in italics. All

TABLE III.—PRODUCT-MOMENT CORRELATIONS BETWEEN INVENTORY SCORES AND STAR RATINGS BY COMPOSITE JUDGMENT OF ADMINISTRATORS AND COUNSELORS

Inventory	Boys		Girls		Total	
	N	*r*	*N*	*r*	*N*	*r*
Sims Score Card for Socio-Economic Status	179	*−.23 ± .05*	172	−.11 ± .05	351	*−.18 ± .03*
Bell Adjustment Inventory						
a. Home	181	.11 ± .05	192	*.25 ± .05*	373	*.15 ± .03*
b. Health	181	.11 ± .05	192	.14 ± .05	373	*.13 ± .03*
c. Social	181	.15 ± .05	192	.03 ± .05	373	.08 ± .03
d. Emotional	181	.16 ± .05	192	.11 ± .05	373	.11 ± .03
Total score	181	.18 ± .05	192	.18 ± .05	373	*.17 ± .03*
Bernreuter Personality Inventory						
N. Neurotic tendency	169	.11 ± .05	161	.12 ± .05	330	.10 ± .04
S. Self-sufficiency	169	−.07 ± .05	161	.13 ± .05	330	.02 ± .04
D. Dominance	169	−.05 ± .05	161	.01 ± .05	330	−.02 ± .04
Loofbourow-Keys Personal Index						
Questionnaire only	179	*.28 ± .05*	196	*.20 ± .05*	375	*.24 ± .03*
Total by 1933 edition	179	*.28 ± .05*	196	*.39 ± .04*	375	*.33 ± .03*
Total by 1937 edition	179	*.31 ± .05*	196	*.41 ± .04*	375	*.35 ± .03*

correlations involving the Personal Index are positive and significant, though lower than might have been anticipated. Of the one hundred eight other coefficients, however, but twenty-one are large enough to afford this practical certainty that the true relationship is in the direction indicated by the sign.

THE UNRELIABILITY OF BEHAVIOR RATINGS

A plausible explanation of the low correlations discovered would be lack of reliability or validity in the criteria of behavior. The best

measure of the reliability of ratings such as here employed is the correlation between ratings assigned by different judges or groups of judges. The present study afforded no means of determining this for H-O-W and star ratings separately. It was, however, possible to correlate the two sets of ratings with one another. This procedure showed for the boys an r of .26 ± .05, for the girls, .55 ± .03, and for the total population, .38 ± .03.

TABLE IV.—MODIFIED BI-SERIAL CORRELATIONS BETWEEN INVENTORY SCORES AND STAR RATINGS BY COMPOSITE JUDGMENT OF ADMINISTRATORS AND COUNSELORS

Inventory	Boys		Girls		Total	
	N	r_{bis}	*N*	r_{bis}	*N*	r_{bis}
Sims Score Card for Socio-Economic Status	179	−.23 ± .07	172	.08 ± .07	351	−.18 ± .04
Bell Adjustment Inventory						
a. Home	181	.24 ± .06	192	.40 ± .06	373	.32 ± .04
b. Health	181	.29 ± .06	192	.27 ± .06	373	.28 ± .04
c. Social	181	.05 ± .06	192	−.12 ± .06	373	−.05 ± .04
d. Emotional	181	.29 ± .07	192	.10 ± .06	373	.14 ± .04
Total score	181	.30 ± .06	192	.20 ± .06	373	.21 ± .04
Bernreuter Personality Inventory						
N. Neurotic tendency	169	.01 ± .06	161	.02 ± .07	330	.01 ± .05
S. Self-sufficiency	169	.12 ± .06	161	.23 ± .07	330	.17 ± .05
D. Dominance	169	.14 ± .06	161	.13 ± .07	330	.17 ± .05
Loofbourow-Keys Personal Index						
Questionnaire only	179	.51 ± .05	196	.74 ± .04	375	.60 ± .03
Total by 1933 edition	179	.51 ± .05	196	.61 ± .05	375	.55 ± .04
Total by 1937 edition	179	.59 ± .05	196	.64 ± .05	375	.60 ± .03

The variables in the situation are too numerous to warrant explanations as to why homeroom teachers agreed so much better with the administrative group in their ratings of girls than of boys. That they did so, is unmistakable. It must be remembered, however, that the same teachers did not judge both sexes; and that, whereas teachers of girls' homerooms were all women, not all of the boys' groups were rated by men.

More important is the fact that the three coefficients average but .40. While the reliability of the star ratings was probably higher

than this and that of the H-O-W ratings somewhat lower, the average reliability was apparently of the order of .4.

CONCLUSIONS: DEGREE OF PREDICTION OBTAINABLE

This information throws fresh light on the correlations of Tables I to IV. If, for example, the star ratings be accepted as the best criterion of problem behavior here available, it is evident that the Haggerty-Olson-Wickman ratings agreed with these to the extent of only .38 for the total group. The Personal Index in its 1937 form, however, shows a product-moment correlation with star ratings of .35, and a bi-serial *r* of .60. This test, therefore, serves to predict behavior as judged by a group of administrators and counselors as well as do painstaking ratings by instructors who have been for from two to five semesters the home-room teachers of the pupils concerned.

In Table V the preceding data are presented in summary form. The twelve coefficients entering into each average in the next to the last column have been weighted inversely as the square of their probable errors.[1] Among other advantages, this serves to give greater weight to those correlations based on a larger number of cases. In the final column, the mean coefficients are shown corrected for attenuation on the assumption of an average reliability of .40 for the ratings, and reliabilities of the inventories or parts thereof equal to those reported in the respective authors' manuals. These corrected coefficients are presented for the reader's convenience, and should be viewed as approximations only.

Inspection of this and preceding tables leads to the following conclusions:

1. For no test or inventory are the correlations shown sufficient for accurate prediction of the behavior of individual pupils. Scores, even on the Personal Index, must be regarded as suggestive only.

2. Correlations based on the Personal Index are all significant, and conspicuously higher than those for other inventories. For the 1937 form, these average .41 before, and .67 after correction for attenuation. While much lower than those obtained by the authors on original standardization, and disappointing in an instrument specifically designed to identify disciplinary problem cases, these might still be held to indicate "a degree of prediction comparable to that with

[1] Kelley, T. L.: *op. cit.*, pp. 324–325, "The Best Weighted Average of Independent Variables."

which a good intelligence test forecasts achievement in school subjects."[1]

TABLE V.—RANGE AND MEAN OF TWELVE COEFFICIENTS OF CORRELATION FOR EACH INVENTORY

Inventory	Number of correlations statistically significant	Range of *r*'s	Mean *r*	
			Uncorrected	Corrected for attenuation
Sims Score Card for Socio-Economic Status	3	−.23 to +.21	$-.09 \pm .02$	−.15
Bell Adjustment Inventory				
a. Home	8	+.03 to +.40	$.20 \pm .02$	.34
b. Health	4	−.02 to +.29	$.13 \pm .02$	.23
c. Social	0	−.24 to +.15	$-.02 \pm .02$	−.03
d. Emotional	1	−.06 to +.29	$.08 \pm .02$	.14
Total score	3	−.14 to +.30	$.12 \pm .02$	.20
Bernreuter Personality Inventory				
N. Neurotic tendency	0	−.12 to +.12	$.01 \pm .01$	.02
S. Self-sufficiency	0	−.08 to +.23	$.07 \pm .02$	.12
D. Dominance	2	−.05 to +.32	$.12 \pm .02$	.20
Loofbourow-Keys Personal Index				
Questionnaire only	12	+.20 to +.74	$.38 \pm .03$	.62
Total by 1933 edition	12	+.23 to +.61	$.37 \pm .02$	.61
Total by 1937 edition	12	+.26 to +.64	$.41 \pm .02$	.67

Lewin,[2] for example, in a somewhat similar situation, has recently reported an average correlation of but .40 between MA on the Terman Group Test and grades in five first-year subjects of three San Francisco high schools. Maller,[3] also, found for five thousand seven hundred sixty-three seniors in five large high schools an average correlation of only .36 ± .01 between high-school grades and Terman Group IQ.

[1] Loofbourow, G. C., and Keys, Noel: "A Group Test of Problem-Behavior Tendencies in Junior-High-School Boys." *Journal of Educational Psychology*, Vol. XXIV, pp. 641–653, December, 1933, p. 651.

[2] Lewin, Lillie: *The Prediction of Success in Certain Academic Subjects in First Year Senior High School.* Unpublished Ph. D. dissertation, University of California, Berkeley, Calif., 1937, p. 70.

[3] Maller, J. B.: "Age *versus* Intelligence as Basis for Prediction of Success in High School." *Teachers College Record*, Vol. XXXIII, February, 1932, pp. 402–415.

3. The 1937 form of the Index shows higher validity than the 1933. Though differences observed are too small for certainty, omission of the "virtues" test was apparently an improvement.

4. The value of the Index seems to lie chiefly in the adjustment questionnaire, the other two sections adding only moderately to the prediction afforded. That this questionnaire should show to better advantage than those of the Bell and Bernreuter inventories is perhaps explainable by the many items dealing with adjustment in school situations. The closer agreement cannot be attributed to knowledge by teachers and administrators of Index scores, since these papers were marked at the University of California and no results reported until all ratings had been submitted.

5. Of the nine other inventories or sections compared, six show mean correlations sufficient for statistical significance, though too small for practical importance.

6. Of these, the highest prediction is afforded by the home adjustment section of the Bell Inventory, with an average correlation of .20 before, and .34 after correction for attenuation. These coefficients exceed those for the Sims Score Card by a statistically significant margin. Evidently the "goodness" of a home from the standpoint of children's behavior is more a matter of human relationships than of economic circumstances.

7. Poor social-economic status as measured by the Sims Card is significantly associated with problem behavior in boys as rated by principal and counselors. This is not true, however, of girls or of ratings by homeroom teachers.

8. Health adjustments (Bell) are found significantly related to problem behavior as rated by administrators and counselors, but not by homeroom teachers.

9. Social and emotional adjustment as measured by the Bell Inventory show very slight relation to disciplinary conduct.

10. Of the Bernreuter scales used, scores for neuroticism and self-sufficiency yield no significant correlation with problem behavior. A tendency is observable, however, for homeroom teachers to rate the more dominant individuals as problems. The mean correlation for this trait is significant, though negligible ($.12 \pm .02$).

SUMMARY

The present study compares the serviceability of twelve tests or subtests of the inventory type as bases for the prediction of disciplinary

problem behavior. Appraisals of such behavior, though obtained under favorable circumstances, proved exceedingly unreliable. Correlations between the ratings of homeroom teachers and the composite judgment of a group of administrators and counselors averaged but .40. Of the four inventories compared, only the Loofbourow-Keys Personal Index, which was specifically designed for this purpose, afforded predictions sufficiently high to be of practical value for the counselor. Scores on the 1937 form of this test agreed as well with the composite judgment of counselors and administrators as did ratings made with the help of Haggerty-Olson-Wickman Behavior Schedules by instructors who had served as homeroom teachers of the pupils concerned for from two to five semesters. Low but statistically significant correlations with problem-behavior were also found for scores on the Bell Adjustment Inventory and three subsections thereof, the Sims Score Card for Socio-Economic Status, and the dominance-submission scale *B4-D* of the Bernreuter Personality Inventory.

THE MEASUREMENT OF OVERLAPPING

J. W. TILTON

Yale University

Ordinarily, the extent to which one distribution overlaps a second is represented by the percentage of scores in the first which either equal or exceed the median score of the second distribution. This measure is unsatisfactory because it has a variable and undefined relation to the total amount of overlapping. Its maximum value, fifty per cent, describes one hundred per cent overlapping. But when twenty-five per cent of one distribution equals or exceeds the median of the second, (see Figure I) the two distributions overlap to the extent of seventy-four per cent, in the sense that the scores of seventy-four per cent of either group may be matched with scores from the other group. In other words, the shaded area in Figure I, which is the area of overlapping, is seventy-four per cent of the area under either frequency curve. When only nine per cent reach or exceed the median, (see Figure II) still the two groups are as much alike as they are different in the sense that they are alike in fifty per cent of the cases. A fully or adequately measured overlapping of nineteen per cent (see Figure III) would ordinarily be reported as zero overlapping since, under these conditions, less than one-half of one per cent of one distribution might be expected to reach or exceed the median of the other.

Symonds[1] has suggested the use of bi-serial r as a measure of overlapping, after considering, as an alternative, the difference between the means of the distributions. Bi-serial r is a good substitute for the usual method of measuring overlapping, except for the fact that its minimum value .00 is a measure of maximum overlapping and 1.00 indicates less than one-half of one per cent of one distribution equalling or exceeding the mean of the other. In Figure 1 of Symond's article, it is incorrectly represented as superior to the common measure of overlapping from the standpoint of reliability. This is due to the fact that Symonds used as the standard error of the per cent of overlapping, $\sqrt{\frac{.25}{N_1} + \frac{pq}{N_2}}$. The $.25/N_1$ should be multiplied by the ratio of the ordinates of the distributions at the median. When, in the case of two normal distributions of equal populations, the overlapping is complete, the ratio

[1] Symonds, P. M.: "A Comparison of Statistical Measures of Overlapping with Charts for Estimating the Value of Bi-serial r." *Journal of Educational Psychology*, Vol. XXI, pp. 586–596.

omitted equals 1, and its omission introduces no error. But as the overlapping decreases, the ratio approaches zero, and appreciably reduces the standard error, as may be seen in Kelley's[1] Table LXXI. Had Symonds used these more accurate values of the standard error of the per cent of overlapping, bi-serial r would not have been represented as a more reliable measure of overlapping.

The chief limitation of bi-serial r is the limitation, already described, of the common percentage measure of overlapping for which it is offered as a substitute. When it has increased to 1.00 there is still

ORDINARILY REPORTED AS 25%

FIG. I

ORDINARILY REPORTED AS 0%

FIG. III

ORDINARILY REPORTED AS 9%

FIG. II

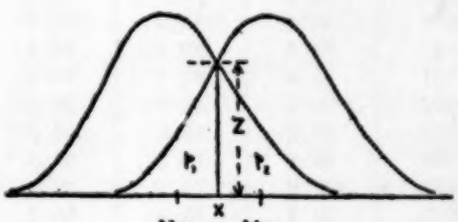

FIG. IV

nineteen per cent overlapping of distribution areas as previously illustrated in Figure III. It is true that the formula can be used when there is less overlapping. When bi-serial r is 1.18 or greater there is less than one-half of one per cent of overlapping area, or, for practical purposes, no overlapping. But the incorrect use of a correlation formula to give values greater than 1.00 does not commend itself.

The difference between means in sigma units is a very good measure of overlapping. With such a measure, a person familiar with the relation of the standard deviation to a normal distribution can reconstruct a very satisfactory picture of the overlapping measured. For the statistically untrained the measure is unsatisfactory. For them it emphasizes only the difference between the groups. Because of this fact, the comparison of means should be supplemented whenever possible by an explicit measure of overlapping, such as the percentage of area common to the two distributions, which is, in other words, the

[1] Kelley, T. L.: *Statistical Method.* New York: The Macmillan Company, 1924, p. 319.

TABLE I.—PERCENTAGE OF OVERLAPPING (% Ov.)

$\left(\text{col. 1, } \frac{Mn - Mn}{SD}\right)$ (col. 2, % = or > Mn) (col. 3, % Ov.) (col. 4, $\sqrt{N}$ SE % Ov.)

1	2	3	4	1	2	3	4
0.000	50.0	100.0	56.4	1.349	8.9	50.0	54.2
0.025	49.0	99.0	56.4	1.381	8.4	49.0	54.0
0.050	48.0	98.0	56.4	1.413	7.9	48.0	53.8
0.075	47.0	97.0	56.4	1.445	7.4	47.0	53.6
0.100	46.0	96.0	56.4	1.478	7.0	46.0	53.4
0.125	45.0	95.0	56.4	1.511	6.5	45.0	53.1
0.151	44.0	94.0	56.4	1.544	6.1	44.0	52.9
0.176	43.0	93.0	56.4	1.578	5.7	43.0	52.6
0.201	42.0	92.0	56.4	1.613	5.3	42.0	52.4
0.226	41.1	91.0	56.4	1.648	5.0	41.0	52.1
0.251	40.1	90.0	56.4	1.683	4.6	40.0	51.7
0.277	39.1	89.0	56.4	1.719	4.3	39.0	51.4
0.302	38.1	88.0	56.4	1.756	4.0	38.0	51.1
0.327	37.2	87.0	56.4	1.793	3.6	37.0	50.7
0.353	36.2	86.0	56.4	1.831	3.4	36.0	50.3
0.378	35.3	85.0	56.4	1.869	3.1	35.0	49.9
0.403	34.4	84.0	56.4	1.908	2.8	34.0	49.5
0.429	33.4	83.0	56.4	1.948	2.6	33.0	49.0
0.455	32.5	82.0	56.4	1.989	2.3	32.0	48.5
0.481	31.5	81.0	56.4	2.030	2.1	31.0	48.0
0.507	30.6	80.0	56.4	2.073	1.9	30.0	47.5
0.533	29.7	79.0	56.3	2.116	1.7	29.0	46.9
0.559	28.8	78.0	56.3	2.161	1.5	28.0	46.3
0.585	27.9	77.0	56.3	2.206	1.4	27.0	45.7
0.611	27.1	76.0	56.3	2.253	1.2	26.0	45.1
0.637	26.2	75.0	56.3	2.301	1.1	25.0	44.4
0.664	25.3	74.0	56.3	2.350	0.9	24.0	43.7
0.690	24.5	73.0	56.2	2.401	0.8	23.0	42.9
0.717	23.7	72.0	56.2	2.453	0.7	22.0	42.1
0.744	22.9	71.0	56.2	2.507	0.6	21.0	41.2
0.771	22.1	70.0	56.2	2.563	0.5	20.0	40.3
0.798	21.3	69.0	56.1	2.621	0.4	19.0	39.4
0.825	20.5	68.0	56.1	2.682	0.4	18.0	38.4
0.852	19.7	67.0	56.0	2.744	0.3	17.0	37.4
0.880	18.9	66.0	56.0	2.810	0.2	16.0	36.3
0.908	18.2	65.0	55.9	2.879	0.2	15.0	35.1
0.935	17.5	64.0	55.8	2.952	0.2	14.0	33.8
0.963	16.8	63.0	55.8	3.028	0.1	13.0	32.5
0.992	16.1	62.0	55.7	3.110	0.1	12.0	31.1
1.020	15.4	61.0	55.6	3.196	0.1	11.0	29.7
1.049	14.7	60.0	55.5	3.290	0.0	10.0	28.1
1.078	14.1	59.0	55.4	3.391	0.0	09.0	26.4
1.107	13.4	58.0	55.3	3.501	0.0	08.0	24.6
1.136	12.8	57.0	55.2	3.624	0.0	07.0	22.6
1.166	12.2	56.0	55.1	3.762	0.0	06.0	20.5
1.196	11.6	55.0	55.0	3.920	0.0	05.0	18.2
1.226	11.0	54.0	54.9	4.107	0.0	04.0	15.6
1.256	10.5	53.0	54.7	4.340	0.0	03.0	12.8
1.287	09.9	52.0	54.6	4.653	0.0	02.0	09.5
1.318	09.4	51.0	54.4	5.152	0.0	01.0	05.7

percentage of scores made by one group which could be matched with scores in the other group.

This full measure of overlapping is often the one wanted in educational comparisons. Starch[1] used it to describe the overlapping of school grades. Burr[2] could have used it to very great advantage to supplement his graphic portrayal of the overlapping among "homogeneous" sections instead of the ratio of section to grade range. Terman and Miles[3] used a measure which is equal to one-half of the total percentage of overlapping.

Assuming the need for such a measure, the writer suggests that it be obtained as for normal distributions separated by the number of sigma scores between the means of the overlapping distributions. For two overlapping normal distributions of equal N and equal sigma, as in Figure IV, the intersection lies midway between the means. The percentage area on either side of "Z" at the point of intersection, *i.e.*, of p_1 or p_2, may be read from a table of the normal probability integral. The desired entry is one-half the sigma distance between the means. Twice this area, *i.e.*, $p_1 + p_2$, is the full measure of overlapping desired.

Table I[4] has been prepared to serve three purposes: To save the reader the necessity of halving the sigma distance between the means, and doubling the integral found in a probability table; to facilitate the computation of standard errors (column four); and to show the relation between the full measure of overlapping (column three), the customary measure (column two), and the sigma distance between means (column one). In other words, for each per cent of overlapping, in column three, is given: in column one, the difference between means in terms of sigma for which it is appropriate; in column two, the corresponding measure of overlapping as it has been commonly defined; and in column four, the $\sqrt{N}$ times the standard error of the desired percentage found in column three.

Since the table is based upon the assumption of normality, and of equality of N and sigma, the desirability of its use is certain for only those situations in which departure from normality and from equality

---

[1] Starch, D.: *Educational Psychology.* New York: The Macmillan Company, 1926, p. 36. (And in the 1928 revision, on p. 37.)

[2] Burr, M. Y.: *A Study of Homogeneous Grouping.* (Contributions to Education, No. 457.) New York: Teachers College, Columbia University, Bureau of Publications, 1931.

[3] Terman, L. M., and Miles, C. C.: *Sex and Personality.* New York: McGraw-Hill Book Company, Inc., 1936, p. 67.

[4] The Kelley Wood Table was used in the preparation of Table I. Kelley, T. L.: *Op. cit.*, pp. 373–385.

of N and of sigma are not significant. In expressing the difference between means as a sigma distance, the use of the average sigma is recommended. It may be necessary to modify the procedure to fit other situations, but because of the simplicity of the method it is to be hoped that it will prove satisfactory in most situations. The fact that the proposed percentage does not exactly equal the percentage found by the method of counting, is for most purposes an advantage rather than a disadvantage. For theoretical purposes, it is frequently more desirable to know what the overlapping would be for normal distributions of the same or equal populations, than to know by counting what it is in the actual situation.

In those cases where means and sigmas (or Medians and Q's) are already computed, the ease of computation is obvious. This fact makes the method very useful, for there are many cases of reported data to which it may be applied with profit. For instance among the valuable data presented by Byrns and Henmon[1] are the facts that six thousand twenty-eight high-school seniors whose parents were in professional occupations made a median scholastic aptitude score of 68.5, Q equal to 23.5, and that thirteen thousand whose parents were unskilled laborers made a median score of 45.8, Q equal to 24.5. The average Q, 24.0, is equivalent to a sigma of 35.5. The difference between the medians is 22.7 or .639 of a sigma. Table I shows, for .639 the interpolated value 74.9 as the per cent of overlapping. The average N = nine thouasnd five hundred fourteen and the square root of this N is 97.5. The column four value for 74.9 is 56.3, which, divided by 97.5, gives .6 of one per cent as the standard error of the 74.9 per cent. In this case, the authors show by their conclusions that they are aware of this large degree of overlapping, but, without a measure, readers often fail to realize it and authors may even prevent a realization of similarity in groups by an emphasis upon the difference between them. For instance, Pintner[2] quotes army testing data from the Memoirs[3] to show the "marked" difference between those from northern Europe and those from southern Europe. It isn't that the difference is not significant, but that it is less marked when viewed with the fact

---

[1] Byrns, R., and Henmon, V. A. C.: "Parental Occupation and Mental Ability." *Journal of Educational Psychology*, Vol. XXVII, pp. 284–291.

[2] Pintner, R.: *Intelligence Testing.* Revised Edition. New York: Henry Holt and Company, 1931, pp. 460 and 467.

[3] Yerkes, R. M., *Editor:* "Psychological Examining in the United States Army." *Memoirs of the National Academy of Sciences*, Vol. XV. Washington: Government Printing Office, 1921, p. 699, Table 218.

that between the Italian group on the one hand, and the Denmark, Norway and Sweden group on the other, there was a 65.3 per cent overlapping.

Another good feature of the method of computing the per cent of overlapping proposed in this article is the ease with which correction may be made for differences in reliability as suggested by Kelley.[1] Suppose two groups are tested with three tests—a completion, an artificial language, and an analogies test—in order to find out in which trait the groups are most alike and in which trait most different. For illustration, means and sigmas may be taken from Livesay's[2] comparisons of Caucasian and Part-Hawaiian groups. At face value they supply the answer that the groups overlap most on the artificial language test and least on the analogies test, the percentages being 87.2, 75.5, and 75.2. Since the reliabilities are given, the true sigmas may be computed by the use of the formula:

$$\sigma_\infty = \sigma\sqrt{r}^*$$

The differences between the means in terms of the true sigmas are greater and the overlappings less. The 87.2 becomes 87.1, 75.5 becomes 72.9, and 75.2 becomes 73.4. In terms of true overlapping the fact remains that the groups differ least in their artificial language scores, but it now appears that they differ most in completion scores, instead of in analogies scores. In this case the effect of the correction is slight because the reliabilities were all high, but it will be evident that in such problems the use of true sigmas may make significant differences in the amount of overlapping.

The standard error of the suggested percentage of overlapping is derived as follows: Let the distributions of Figure IV be normal and their populations equal to 1. The overlapping (Ov.) is the area p_1 plus p_2 or, since $p_1 = p_2$, Ov. $= 2p$. SD $2p$ or 2 SD p is then the SE desired. If it were being suggested that Ov. be obtained by a counting method, its error would be due in part to the SE of a per cent, $\sqrt{pq/n}$, and due in part to the error in the location of the ordinate separating p_1 from p_2. However, the counting method is not recommended. It is suggested rather that the overlapping reported be that for two perfectly

[1] Kelley, T. L.: "Measurement of Overlapping." *Journal of Educational Psychology*, Vol. X, pp. 229–232.

[2] Livesay, T. M.: "Racial Comparisons in Performance on the American Council Psychological Examination." *Journal of Educational Psychology*, Vol. XXVII, pp. 631–634. Table I, p. 632.

* Kelley, T. L.: *Statistical Method*, number 166, p. 213.

normal distributions similarly separated. Its error, therefore, is due solely to the location of the point x. Assuming the distributions to be flat at the point of intersection,[1] the change in p_1 or p_2 produced by a change in x, dx, is $z\,dx$. SE p is then SD $z\,dx$ or z SD x.

$$x = \frac{Mn_1 - Mn_2}{2\,\mathrm{SD}}$$

For an SD of a ratio between two variables, Yule's[2] formula is:

$$\mathrm{SD}^2\frac{x}{y} = \frac{Mn_x}{Mn_y}\left(v_x^2 - 2r_{xy}v_xv_y + v_y^2\right)$$

By substitution, assuming that r between numerator and denominator = 0, and that r between measures in distribution 1 and distribution 2 = 0, and by using the obtained numerator and denominator as means of the same,

$$\mathrm{SE}\frac{Mn_1 - Mn_2}{2\,\mathrm{SD}} = \frac{1}{2\sqrt{2N}}\sqrt{4 + \left(\frac{Mn_1 - Mn_2}{\mathrm{SD}}\right)^2}$$

$$\mathrm{SE}\,p = \frac{z}{2\sqrt{2N}}\sqrt{4 + \left(\frac{Mn_1 - Mn_2}{\mathrm{SD}}\right)^2}$$

$$\mathrm{SE\ Ov.} = 2\,\mathrm{SE}\,p = \frac{z}{\sqrt{2N}}\sqrt{4 + \left(\frac{Mn_1 - Mn_2}{\mathrm{SD}}\right)^2}$$

The $\sqrt{N}$ times this value is listed in column four of Table I. In the first line, for one hundred per cent it is 56.4. For fifty per cent overlapping, as usually defined, the corresponding value in Kelley's Table LXXI[3] is 70.7. In his table, 10.5 is the value for one per cent. The corresponding per cent in Table I is 24.5 per cent and its $\sqrt{N}\cdot$SE is 44.0. In other words, the standard errors given in Table I are smaller for large amounts of overlapping and larger for the lesser degrees of overlapping. When the measures of overlapping are expressed as multiples of their standard errors, the proposed percentage is more reliable throughout the range of the measure ordinarily used. In this relative sense, the complete measure of overlapping proposed is 1.8 times as reliable as the ordinary percentage in the case of one hundred per cent overlapping, and thirteen times as reliable for the smallest amount of overlapping listed in Kelley's Table LXXI.

---

[1] Following Kelley, T. L.: *Statistical Method*, p. 317.

[2] Yule, G. U.: *An Introduction to the Theory of Statistics*, London: Charles Griffin and Company, Limited, 1922, p. 215, No. 10.

[3] Kelley, T. L.: *Statistical Method*, p. 319.

AN INVESTIGATION OF THE VOCABULARY PROBLEM IN COLLEGE BIOLOGY

ELMO N. STEVENSON

Eastern Oregon Normal School

Is there a vocabulary problem in the study of general biology in college? Many observant, efficient teachers and most of the students of biology believe there is. They are cognizant of the great burden occasioned by the technical nomenclature, associated vernacular, and names of the contributing personnel of this subject.

Research investigations on the junior and senior high-school levels have revealed the magnitude and indicated the possible implications of the vocabulary problem there. Pressey[1] conducted the earliest extensive study of the science vocabulary found in widely used textbooks. The investigator determined an "essential," an "accessory," and an "auxiliary" vocabulary for various high-school sciences and other subjects by using an elaborate technique which employed teacher judgment principally. In commenting upon the length and difficulty of her vocabulary lists she quoted the opinion of foreign-language teachers to the effect that a pupil does well to acquire from eight hundred to a thousand new words in a foreign language; then states that "the pupil in any of the sciences . . . has as great, or a greater, vocabulary to master, wholly aside from acquisition of considerable subject-matter." In the science category, general science led with a total vocabulary of fifteen hundred sixty-four, and biology was next with thirteen hundred ninety-three. She concludes by indicating that the difficulty which children have in the sciences is frequently a meaning difficulty.

Powers,[2] working coincidentally but using a more objective investigational technique, has contributed two extensive studies of the nature and difficulty of the vocabulary found in high-school science textbooks. His first investigation was an analysis of more than a million running words in textbooks of general science, chemistry, and biology, to determine the number and frequency of appearance of uncommon words. He concludes, "It can hardly be denied that the vocabulary

---

[1] Pressey, Luella C.: "The Determination of the Technical Vocabularies of the School Subjects." *School and Society*, Vol. XX, July 19, 1924, pp. 91–96.

[2] Curtis, F. D.: *Second Digest of Investigations in the Teaching of Science*, Vol. II, pp. 348–351.

burden of all science texts is unnecessarily large." He suggests that this vocabulary difficulty be reduced by substitution of phrases for technical terms, and by the introduction of "only such technical terms . . . as the authors find need for after they are introduced."

Powers's second investigation[1] was an extension of his previous study to include an analysis of more texts, scientific books, and periodical articles. His object was to determine "a list of 'uncommon' words found to be most important on the basis of range and frequency of usage in science texts and in other reading material about scientific subjects." He concludes that authors are using in their books many words which children will never see again after leaving the classroom, and that intelligent selection on the part of authors in regard to vocabulary load and difficulty "is of paramount importance in the work of adapting the materials of science textbooks to the ability levels of the pupils for whom they are written." He lists eighteen hundred twenty-eight "uncommon" scientific terms which his analyses have shown to be important and which lie outside of Thorndike's list.

Baird[2] conducted an analysis of student's biology notebooks from fifty representative secondary schools in New York State to determine the degree of agreement in vocabulary used in biology texts and biology notebooks as written by students, and to compare the scientific terms used in notebooks with those found in Powers's List of Scientific Terms for High-School Students and with those in The Teacher's Word Book. His technique consisted of constructing an initial word list against which he checked the vocabulary content of all the notebooks. He found that the total number of separate words used (names and symbols included) was two thousand three hundred fifty-eight. Of these scientific terms, six hundred forty-five were used in all notebooks. Most of them were used but once. He concludes that biology has a technical vocabulary more or less peculiar to it; that the acquirement of this vocabulary is difficult; and that the technical vocabulary of textbooks may well be simplified, especially since many words were not found in Powers's list.

---

[1] Powers, S. R.: "Important Terms Compiled from Textbooks for General Science, Biology, Physics, and Chemistry, and from Other Reading Materials for Scientific Subjects." *Bureau of Publications, Teachers College, Columbia University*, New York, 1926.

[2] Baird, Don O.: "A Vocabulary Study of Biology Notebooks of Fifty Representative Secondary Schools in New York State." *Education*, Jan., 1931, pp. 263–269.

Recent writers of textbooks of high-school science have taken cognizance of the findings of Powers, Baird, Curtis, Pressey, and others. As a result, they are using a simpler vocabulary, a greater simplification of materials of instruction, and are attempting to secure a mastery of essential technical vocabulary through repetition.

The author of this article conducted an investigation at the Eastern Oregon Normal School in general biology to determine whether or not a similar vocabulary load and difficulty existed on the college level. The course in general biology was a year survey course. The principal aims were to develop an understanding and appreciation of the living world; to develop a functional understanding of the major generalizations of biology and the development of associated scientific attitudes; and to provide the student with the practical information that a prospective citizen and teacher needs to adapt himself effectively to his environment. The content consisted of a comprehensive survey of the lower plants and animals, and general principles of biology for the first quarter. The second quarter was devoted to the higher plants and animals, while the third quarter concerned human biology. The method of teaching used was a combination of the problem, lecture-demonstration, discussion, and laboratory procedures. The textbook was *Foundations of Biology* by L. L. Woodruff. The chief references were: *Animal Biology*, by M. F. Guyer, *College Zoology*, by R. W. Hegner, *Textbook of General Botany*, by G. M. Smith and others, *Applied Eugenics*, by P. Popenoe and R. H. Johnson, and *Manual of Biology*, by G. A. Baitsell.

The author's procedure was to construct a master word list from weekly word lists handed in by students. These weekly lists were considered a part of the regular work. All unknown vocabulary items, technical or otherwise, encountered in the course were to be included. The students were informed that the purpose of the study was to improve the course by reducing the excessive vocabulary burden, if there was one. At the end of the quarter the master word list, directions for checking the words, and a request for a few responses to questions were given to each student. Only the directions and beginning of the master list are included here because of space limitations.

Master List

The following new words and phrases were encountered by students in their reading and lectures during the first quarter of a course in general

biology. This list was compiled from word lists handed in weekly. The word or phrase had to be mentioned by at least ten students to be included.

I. Please answer briefly the following questions:
   a. Was the nomenclature of biology a serious handicap to you?
   b. Would you have liked the subject better if there had been fewer new and unfamiliar words and phrases?
   c. What percentage of these words do you think you will ever meet again outside of a biological science course? Underline: 0–25 per cent, 25–50 per cent, 50–75 per cent, 75–100 per cent
   d. Do you plan to take more courses in biological science? If so, do you still think as you have indicated above?

II. Please go over the list and check the words as follows:
   a. Place a check mark (✓) in front of all words which were new or unfamiliar to you at the beginning of the quarter. By unfamiliar is meant that you could not illustrate nor define the word.
   b. Place a cross mark (+) after the words which you think are necessary for an adequate understanding of biology.
   c. Place a minus mark (−) after the vocabulary item which, though not necessary for an adequate understanding of biology, helps to make for better elucidation.

The List

binocular	Reptilia	sulfur	nitrates
algae	Aves	phosphorus	nitrites
diatom	Mammalia	chlorine	radiant energy

These word lists were to be turned in on the final day of the quarter. Students were urged to define or describe as many of the terms as they could. Ample time was allowed for this. Many students conscientiously went through the list as suggested; others started, but gave up for one reason or another.

The master list technique was carried out for the six quarters of the 1933–1934 and 1934–1935 school years. At the end of the fall quarter in 1934 a random sampling check list of twenty-five words to be defined was included in the final test to verify the checking procedure on the master list. The results indicated a high degree of correlation between what had been indicated as known with the degree of achievement on the test list. The reason is evident. This device was used only the one quarter.

The findings are presented in the accompanying tables with sufficient descriptive matter to make them understandable. The statistics have been reduced to the simplest processes. No attempt was made to

TABLE I.—FINDINGS ON WORD FREQUENCIES ON MASTER AND WEEKLY WORD LISTS, AND COMPARISONS WITH AUTHORITATIVE VOCABULARY LISTS

Descriptions	School year					
	Fall		Winter		Spring	
	1933	1934	1934	1935	1934	1935
Vocabulary items on master list (frequency of at least ten in class of thirty basis of list)	1083	743	540	433	313	303
Per cent less than corresponding quarter	.....	31.4	.....	19.8	.....	3.2
Range of checking on master list—low to high.	162–1013	128–687	103–505	55–385	115–312	193–292
Range of different vocabulary items on original weekly list (individual student totals).	213–2153	190–1503	151–1102	101–870	202–513	191–511
Mean of master list checks	685.1	472.5	351.4	309.3	228.6	223.5
Number of cases	66	71	56	63	29	31
Per cent less than corresponding quarter	.....	31.0	.....	12.0	.....	2.2
Master list words in Power's list	243	189	57	44	21	19
Per cent of total on master list	22.4	25.4	10.6	10.2	6.7	6.3
Master list words in Thorndike list	67	23	9	5	3	3
Per cent of total words on master list	6.1	3.1	1.7	1.2	1.0	1.0
Master list words in text glossary	188	197	52	63	20	19
Per cent of total words on master list	17.4	26.5	9.6	14.6	6.4	6.3
Percentage of Woodruff (text) glossary	34.75	43.68	11.529	13.968	4.434	4.212
Definable or describable terms at end of quarter. Mean	598	570	381	397	222	283
Number of cases	21	11	19	21	16	7
Percentage of total master list	55.2	76.7	70.6	91.7	70.9	93.4

determine standard errors. Such procedures seemed to be unnecessary in view of the nature of the problem, the techniques used, and the purposes of the author in making the study.

Several interpretations and generalizations may be drawn from the data presented in Table I. The vocabulary burden of general biology is tremendous. The greatest number of words listed by any student for the 1933–1934 school year was three thousand three hundred twenty-one. The accumulative mean for the corresponding period as revealed by the master word list check was twelve hundred sixty-five. The total listed on the master list was nineteen hundred thirty-nine. (Each of these words was listed by ten or more students.) Over a thousand words were contacted in the first quarter, half that number in the second quarter, and a third in the last quarter.

Comparisons with authoritative vocabulary lists for elementary, secondary, and college students (text glossary) indicated that fifty to eighty-five per cent of the terms on the master list were not included. The comparison with Powers's list, which is primarily a utilitarian and common frequency list, indicated that a large per cent of the words contacted would have little value to the student in everyday life. The data suggested that the greater the vocabulary burden, the smaller the per cent of knowable terms there will be. Only fifty-five per cent of the words in the fall 1933 master list were definable and describable by the students, whereas seventy-five per cent of the fall 1934 master list, with its fewer words by thirty-one per cent over the 1933 list, were definable and describable.

The data reveal a wide difference in vocabulary burden in the group of students concerned. Some have an excellent foundation of science vocabulary; others a poor one; and most of the students need a greater background of the vocabulary that appears to be peculiar to general biology for the best comprehension. The vocabulary burden can be cut down by using phrases and descriptions rather than technical words, and by eliminating nomenclature that is not met in life situations. If this is done, the student takes an increased interest in the subject, and the course becomes more popular. This is desirable for the student who is not specializing in science. Even in his case there should be ample provision for increased vocabulary and learning, but not to the point of overloading or of losing his interest.

A check was made in a German class in the same institution on vocabulary items, using much the same technique. The vocabulary burden for the first quarter was nearly four hundred words, not includ-

ing different forms of the same verb, and for the second quarter nearly eight hundred. The student of biology was expected to learn principles and facts of a subject field and develop attitudes and skills therein in addition to mastering a vocabulary as great as, and in many cases greater than, the student studying a foreign language. The progression is from the smaller to the larger number of words in the language. In biology the greatest vocabulary burden comes early in the course. Obviously, the difference in the nature of the vocabulary factors in the different subject fields must be recognized; *i.e.*, similarity of symbols with known languages, continuity and development of meanings, functional value, etc.

As a further check on the vocabulary burden, the author audited a course in music appreciation and listed the words that were unknown

TABLE II.—STUDENT OPINION ON THE VOCABULARY FEATURE OF GENERAL BIOLOGY IN NORMAL SCHOOL AS DETERMINED BY RESPONSES TO QUESTIONS ON A MASTER LIST

Questions	Response	School year			
		1933–1934		1934–1935	
		Number	Per cent	Number	Per cent
Was the nomenclature of biology a serious handicap to you?	Yes.....	128	60.66	61	56.48
	No.....	83	39.33	47	43.51
Would you have liked the subject better if there had been fewer new and unfamiliar words, names, and phrases?	Yes.....	149	70.95	72	67.92
	No.....	61	29.04	34	32.07
Do you plan to take more courses in biological science?	Yes.....	122	60.69	83	82.17
	No.....	79	39.30	18	17.82
If so, do you still think as you have indicated?	Yes.....	126	91.97	75	90.36
	No.....	11	8.02	8	9.63
	Percentage				
What percentage of these words do you think you will ever meet again outside of a biological science course? Underline, 0–25 per cent, 25–50 per cent, 50–75 per cent, 75–100 per cent.	0– 25	96	46.60	31	30.09
	25– 50	83	40.29	47	45.63
	50– 75	24	11.65	25	24.27
	75–100	3	1.45	0	00.00

to him. (He had escaped most of the contacts with music usually afforded a student.) There were three hundred twenty-one new vocabulary items enumerated. In a graduate course in zoology at the Hopkins Marine Station the author, with a biology minor, checked three hundred forty-three unfamiliar terms during one quarter. During a quarter course in forest botany and distribution he learned more than three hundred new terms. There are few courses other than those in beginning science, especially the survey types, that have an excessive vocabulary impediment.

Table II gives the data on student's responses to questions asked on the master list. A majority percentage say that the nomenclature is a serious handicap. About seventy per cent would like the subject much better if the vocabulary burden were lessened. Seventy-five to eighty per cent of the students expect to meet fifty per cent or less of the words outside of science courses. Even though students contemplate further course work in science, they still feel the same.

A superficial attempt was made to determine whether there was a significant difference in vocabulary burden of students in the different quartiles, based on the Thurstone Psychological Test. Table III presents the results.

TABLE III.—MEANS FOR THE TOTAL WORDS LISTED DURING THE 1933–1934 SCHOOL YEAR BY TEN STUDENTS (RANDOM SAMPLE) IN EACH OF THE QUARTILE GROUPINGS. QUARTILE FOUR IS THE HIGHER ONE

	Quartile			
	I	II	III	IV
Mean	1270.9	1270.2	1120.2	1275.4

The difference between the first, second, and fourth quartiles is insignificant. The difference between the third and fourth is the greatest. Apparently, the vocabulary burden is equally great for all levels of student intelligence. Further study should be made on this phase of the problem before any definite conclusions may be drawn.

The means of the 1933–1934 and 1934–1935 populations as based on the Thurstone Psychological Examination showed a difference of five points. This would indicate that the groups were comparable. The comparative scores on the Coöperative General Biology Test for 1933 and 1934 were one hundred five and one hundred ten respectively. These tests were given at the end of the school year. The population

with the reduced vocabulary burden performed best. The same form was given for both school years.

The data for the fall quarter of the 1933–1934 school year indicated that those students who came from the large towns and had taken biology had shorter word lists than those from village schools. The mean difference was one hundred sixty-seven words. Those who had taken biological science in high school had shorter lists than those who had not. Generally speaking, the poorer students' (class performance) word lists were briefer than those of the better students. This might be due to procrastination, functioning of defense mechanisms, lack of coercion, etc.

The master lists of those students (seventy-three) who completed the 1933–1934 school year were analyzed to determine which words were checked most frequently as being new and unfamiliar to them at the beginning of each quarter. Slightly more than six per cent of the first quarter's words received a unanimous vote, as did nine per cent of the second and eleven per cent of the third quarter's words. Less than three per cent of these new and unfamiliar words are found in Powers's list. This evidence indicates that the number of new and unfamiliar words common to all increases progressively with the study of biology, and that a little over ninety-seven per cent of these new words have slight functional value.

The same selected lists were used for an analysis of the words checked by the students as necessary for an adequate understanding of biology, or as being not necessary but helping to make for better elucidation. The results indicated that the students were good judges. Nearly eighty per cent of the words checked by seventy-five per cent of the students as being essential were found in Powers's list. Only a small percentage of the words checked as not necessary but helping to make for better understanding of the subject was found in Powers's list. Teachers of biology will agree with these judgments.

In the form of concluding generalizations, the author wishes to offer the following:

1. The vocabulary burden in general biology on the college level seems to be enormous, just as it is on the high-school level.
2. The intensity of the burden varies with the individual student.
3. Understanding, appreciation, achievement, and interest function best under a reduced vocabulary impediment.
4. A large percentage of the biological terminology is low in functional value.

5. The authors of college texts for first-year students could reduce the vocabulary burden by substituting phrases and descriptions for technical words.

6. Instructors of elementary and orientation courses will secure more interest and better performance or understanding if they consider the vocabulary item in class, laboratory, and outside reading assignments. Too often the scholarly performance of the professor is on display. Understanding and interest on the part of the student are far worthier goals.

7. There are psychological aspects of the problem which inhibit the student's best performance and should receive attention through further research and philosophical treatment.

8. The author apologetically presents the above, knowing full well the weaknesses of his techniques and the flimsy grounds upon which he dares to present his conclusions. He hopes that some of the techniques presented may be usable, perhaps modified, for subsequent, more scholarly and accurate experiments in the vocabulary problem.

THE USE OF SUBCORRELATION IN DETERMINING THE PREDICTIVE POWER OF HIGH-SCHOOL GRADES

P. S. DWYER

University of Michigan

Many studies have dealt with the correlation existing between high-school record and first semester academic success. Quite complete summaries have been given by such authors as Segel[1] and Wagner.[2] They seem to agree that the bulk of the coefficients of correlation range from about .30 to about .70 with the average about .55. A study of the results leads one to surmise that the predictive power is better for more homogeneous groups. Thus, for example, the predictive power tends to be relatively high at the urban university which draws the major portion of the student body from the city in which it is located. Degree of homogeneity as to age, sex, etc., also appears to influence the size of the coefficient.

This fact leads to the suggestion, and it is by no means a new one, that better predictive power may sometimes be obtained by making subdivisions of the student records into less heterogeneous groups. This alternative to the method of partial correlation is hardly applicable to the small college, but it can be applied to the large university. A fairly recent study using this method is that of M. E. Wagner and E. Strabel at the University of Buffalo. For a description of the groups chosen and the results, see their article in *School and Society*.[3]

The technique used in this paper is not homogeneous grouping in the sense previously used, although the objective, the separation into groups of comparable elements, is similar to it. It is here proposed to break up the distribution into subgroups which are approximately linear and homoschedastic, since linearity and homoschedasticity are necessary if r is to measure uniform predictive power. The correlation coefficients for these linear and homoschedastic subgroups are to be known as subcorrelations.

---

[1] Segel, David: *Prediction of Success in College.* Office of Education Department of Interior, Washington, D. C., Bulletin 1934, No. 15, pp. 69.

[2] Wagner, M. E.: "A Survey of the Literature on College Performance Prediction." Chap. 13 of *Studies in Articulation of High School and College.* E. S. Jones, *Editor,* University of Buffalo Studies, Vol. IX. See pp. 194–196.

[3] Wagner, M. E. and Strabel, E.: "Homogeneous Grouping as a Technique for Improving Prediction Coefficients." *School and Society,* Vol. XL, 1934, pp. 887–888.

The files of the office of Educational Investigations of the University of Michigan supplied information dealing with high-school grades and freshman first semester grades for those students entering the School of Literature, Science, and the Arts in 1926. The high-school record is used as the y variable and the college record is used as the x variable in Table I. The grades are grouped in one-half unit intervals which are characterized by class marks. Entries are made for each row under the headings f, Σfd_x, $(\Sigma fd_x)d_y$, fd_y, Σfd_x^2, fd_y^2, and the results summed for all rows which are combined in the subgroup. The subcorrelations are obtained by the use of the formula

$$r(R{:}k_1, k_2, \cdots k_s) = \frac{f\Sigma fd_xd_y - (\Sigma fd_x)(\Sigma fd_y)}{\sqrt{[f\Sigma fd_x^2 - (\Sigma fd_x)^2][f\Sigma fd_y^2 - (\Sigma fd_y)^2]}}$$

where $k_1, k_2, \ldots k_s$ are the class marks of the rows combined.

The application of the test for straightness of regression line[1] does not definitely establish the significance of deviations from the straight line, though a study of a large number of similar charts reveals the consistency of these deviations. In addition the charts show a moderate degree of heteroschedasticity with the top rows having somewhat smaller variances.

Further study of these charts shows that the conditions of linearity and homoschedasticity are more nearly satisfied when the rows having class marks of 8, 7, 6 are selected, when those having class marks of 5, 4 are selected, and when those having class marks of 3, 2, 1 are selected. The detailed calculations of the value of r for these subgroups are shown in the lower righthand corner of Table I.

The first of the triple entries is a three- or four-place approximation to $f\Sigma fd_xd_y - (\Sigma fd_x)(\Sigma fd_y)$. The two entries below are corresponding approximations to $f\Sigma fd_x^2 - (\Sigma fd_x)^2$ and $f\Sigma fd_y^2 - (\Sigma fd_y)^2$, respectively. The first entry in the next column is the subcorrelation.

The values are then:

$$\begin{aligned} r &= .441 \\ r(R{:}8, 7, 6) &= .288 \\ r(R{:}5, 4) &= .156 \\ r(R{:}3, 2) &= .007 \end{aligned}$$

These values are not comparable measures of predictive power. Comparable measures might be derived by computing the coefficient of

[1] Fisher, R. A.: *Statistical Methods for Research Workers.* 5th edition, pp. 239–243.

TABLE I.—HIGH-SCHOOL GRADES *vs.* FIRST SEMESTER AVERAGE, COLLEGE OF LITERATURE, SCIENCE AND THE ARTS, UNIVERSITY OF MICHIGAN, 1926

First Semester Grade

Grade									.00–.49	.50–.99	1.00–1.49	1.50–1.99	2.00–2.49	2.50–2.99	3.00–3.49	3.50–3.99	4.00–
High school	Class mark								0	1	2	3	4	5	6	7	8
		f							35	60	121	179	341	220	107	50	13
			d	Σfd_x	$(\Sigma fd_x)d_y$	fd_y	Σfd_x^2	fd_y^2	−4	−3	−2	−1	0	1	2	3	4
4.00–	8	18	3	41	123	54	123	162					1	5	5	2	5
3.50–3.99	7	106	2	136	272	212	378	424	1	1		7	20	27	29	19	2
3.00–3.49	6	178	1	118	118	178	468	178	2	5	6	18	44	53	35	12	3
2.50–2.99	5	270	0	−15	0	0	641	0	8	11	27	33	103	55	20	10	3
2.00–2.49	4	330	−1	−174	174	−330	782	330	13	19	46	67	114	54	11	6	
1.50–1.99	3	173	−2	−162	324	−346	502	692	7	18	34	44	45	19	5	1	
1.00–1.49	2	51	−3	−49	147	−153	175	459	4	6	8	10	14	7	2		
All.....		1126		−105	1158	−385	3069	2245	1263 3445 2380	.441 .194 .806	2.726 −.093 2.717	1.994 −.342 1.877	1.000 .806 1.000	1.000 .441	2.190		
8, 7, 6 (*A*, *B*).....		302		295	513	444	969	764	239 2056 336	.288 .083 .917	3.209 .977 2.254	2.530 1.470 .369	.197 .181 .264	.514 .560	2.067		
5, 4 (*C*).....		600		−189	174	−330	1423	330	420 8181 891	.156 .024 .976	2.372 −.315 2.273	.550 −.550 .248	.132 .129 .153	.391 .399	2.218		
3, 2 (below *C*).....		224		−211	471	−499	677	1151	2.15 1071 88	.007 .000 1.000	3.022 −.942 2.135	5.138 −2.228 .174	.093 .093 .093	.305 .023	2.135		

a distribution whose y standard deviation is equal to the y standard deviation of the original distribution and yet whose linearity and schedasticity are identical with those of the subdistribution. These measures, $r'(R:k_1, \ldots k_s)$, are to be called corrected subcorrelations and are computed from the formula

$$r'(R:k_1, \cdots k_s) = \frac{r(R:k_1, \cdots k_s)}{\sqrt{c_2^2(R:k_1, \cdots k_s)[1 - r^2(R:k_1, \cdots k_s)] + r^2(R:k_1, \cdots k_s)}}$$

where $c_2^2(R:k_1, \ldots k_s)$ is the ratio of the y squared standard deviation of the subgroup to the y squared standard deviation of the entire distribution.[1]

An alternative measure is the squared standard deviation of errors of the subgroup. It is given by the formula

$$S_1^2(R:k_1, \cdots k_s) = \sigma_1^2(R:k_1, \cdots k_s)[1 - r^2(R:k_1, \cdots k_s)]$$

The further entries in Table I show the computations necessary to determine these measures. The entries under $r(R:k_1, \ldots k_s)$ are $r^2(R:k_1, \cdots k_s)$ and $1 - r^2(R:k_1, \cdots k_s)$, respectively. The entries in the next two columns are $\frac{\Sigma f d_x^2}{n}$, $\frac{\Sigma f d_x}{n}$, $\sigma_1^2 = \frac{\Sigma f d_x^2}{n} - \left(\frac{\Sigma f d_x}{n}\right)^2$ and $\frac{\Sigma f d_y^2}{n}$, $\frac{\Sigma f d_y}{n}$, $\sigma_2^2 = \frac{\Sigma f d_y^2}{n} - \left(\frac{\Sigma f d_y}{n}\right)^2$, respectively. The entries in the next column are $c_2^2(R:k_1, \cdots k_s)$, $c_2^2(R:k_1, \cdots k_s)[1 - r^2(R:k_1, \cdots k_s)]$ and

$$c_2^2(R:k_1, \cdots k_s)[1 - r^2(R:k_1, \cdots k_s)] + r^2(R:k_1, \cdots k_s),$$

respectively. The entries in the next column are

$$\sqrt{c_2^2(R:k_1, \cdots k_s)[1 - r^2(R:k_1, \cdots k_s)] + r^2(R:k_1, \cdots k_s)}$$

and $r'(R:k_1, \ldots k_s)$, respectively. The entry in the last column is $S^2(R:k_1, \ldots k_s)$.

The results of Table I are given in Table II.

It appears that there is not much range in the standard deviation of errors although the top group has a slightly smaller value than the others, but there is a decided range in the size of the corrected subcorrelations.

---

[1] Compare Kelley *Statistical Method*, p. 225.

TABLE II.—SUBCORRELATIONS BETWEEN HIGH-SCHOOL GRADES AND FIRST SEMESTER RECORD FOR COLLEGE OF LITERATURE, SCIENCE, AND THE ARTS, UNIVERSITY OF MICHIGAN, 1926

Group (high school)	f	r	r'	S_1^2
All	1126	.44	.44	2.19
8, 7, 6 (*A*, *B*)	302	.29	.56	2.07
5, 4 (*C*)	600	.16	.40	2.22
3, 2, 1 (below *C*)	224	.01	.02	2.14

The increased predictive power of the top group is largely due to the decreased slope of the line of regression, while the decreased predictive power of the bottom group is due to the increased slope of that line. It should be noted that the study deals only with the predictive power of high-school grades in the case of those students who are admitted to college. The correlation would be much higher if those high-school students who would surely fail were permitted to enter college. It is probable that the low correlation for the low group is due to the fact that those in the lower brackets have been admitted to college by admission officers who felt that there were reasons for discounting the high-school record. The result probably indicates a partial justification of their judgment.

The decreased slope which is revealed by the corrected subcorrelation of the top group, but not by the squared standard deviation of errors, is very important. In problems where the standard deviation of errors is not small, there is little chance of satisfactory prediction, even by groups, if the slope of the line of regression is not small. A small slope enables one to establish correspondence between the different ranges of the x and y scales. None of those making *A* in high school made less than *C* in college.

The question next arises as to the representativeness of this entering class of 1926. Do the classes entering in later years show the same tendency? Do other schools in the University show the same tendency? Do other colleges in Michigan show the same tendency?

The correlation tables for the years 1927–1930 are also available in the Office of Educational Investigations of the University of Michigan. In addition to the correlation coefficients, which the staff had previously computed, the subcorrelations were computed, as in Table I, for all freshmen entering the college of Literature, Science, and the Arts in the years 1926–1930.

TABLE III.—SUBCORRELATIONS BETWEEN HIGH-SCHOOL AVERAGE AND FIRST SEMESTER COLLEGE RECORD FOR FRESHMEN OF THE COLLEGE OF LITERATURE, SCIENCE, AND THE ARTS ENTERING THE UNIVERSITY OF MICHIGAN IN THE YEARS 1926–1930

Group (high school)	f	r	r'	S_1^2
All	4848	.44	.44	2.14
8, 7, 6 (*A*, *B*)	1571	.33	.62	1.89
5, 4 (*C*)	2546	.16	.40	2.17
3, 2, 1 (less than *C*)	731	.02	.05	1.87

It appears that 1926 was a representative year.

The freshmen entering the College of Engineering in the same period constitute another group which can be studied by this technique. The results are presented in Table IV.

TABLE IV.—SUBCORRELATIONS BETWEEN HIGH-SCHOOL AVERAGES AND FIRST SEMESTER RECORD FOR FRESHMEN OF THE COLLEGE OF ENGINEERING ENTERING THE UNIVERSITY OF MICHIGAN IN THE YEARS 1926–1930

Groups (high school)	f	r	r'	S_1^2
All	1275	.42	.42	2.22
8, 7, 6 (*A*, *B*)	349	.33	.63	1.77
5, 4 (*C*)	705	.19	.44	2.35
3, 2, 1 (less than *C*)	221	.07	.22	1.94

The same basic tendencies appear though the low group seems to have better predictive power than indicated above.

The results of an analysis of the records of all students entering the academic schools of the University in 1927–1930 are given in Table V.

TABLE V.—SUBCORRELATIONS BETWEEN HIGH-SCHOOL AVERAGE AND FIRST SEMESTER RECORD FOR FRESHMEN ENTERING THE UNIVERSITY OF MICHIGAN IN THE YEARS 1927–1930

Group (high school)	f	r	r'	S_1^2
All	5238	.43	.43	2.10
8, 7, 6 (*A*, *B*)	1728	.31	.60	1.86
5, 4 (*C*)	2781	.17	.41	2.15
3, 2, 1 (less than *C*)	729	.05	.15	2.19

It appears that the highest preparatory-school group has the best predictive power as indicated by the corrected subcorrelation and by the standard deviation of errors. The predictive power of the average group is about that of the entire distribution as measured by the corrected subcorrelation and by the standard deviation of errors. The lowest group has about average standard deviation of errors, but the increased slope of the line of regression makes a small correlation.

TABLE VI.—SUBCORRELATIONS BETWEEN HIGH-SCHOOL AVERAGE AND FIRST SEMESTER COLLEGE RECORD FOR FRESHMEN ENTERING TWENTY MICHIGAN COLLEGES IN 1928

Group	f	r	r'	S_1^2
All	4951	.49	.49	1.78
8, 7, 6 (*A*, *B*)	1459	.28	.57	1.57
5, 4 (*C*)	2649	.17	.40	1.84
3, 2, 1 (below *C*)	843	.08	.27	1.89

TABLE VII.—CORRECTED SUBCORRELATIONS BETWEEN HIGH-SCHOOL AVERAGE AND FIRST SEMESTER COLLEGE RECORD FOR FRESHMEN ENTERING DIFFERENT MICHIGAN COLLEGES IN 1926–1930

Group	1926 Literature, Science and Arts, University of Michigan	1926–1930 Literature, Science and Arts, University of Michigan	1926–1930 Engineering University of Michigan	1927–1930 totals, University of Michigan	1928 twenty Michigan colleges
All	.44	.44	.42	.43	.49
A, *B*	.56	.62	.63	.60	.57
C	.40	.40	.44	.41	.40
D, *E*	.02	.05	.22	.15	.27

The next question concerns the applicability of this technique to other institutions. In order to show that this same tendency does exist in other schools, at least in the State of Michigan, I have subjected to a corresponding analysis the records of freshmen entering twenty colleges and universities in Michigan in the fall of 1928. A report of a coöperative study was made by Ruth A. Brown,[1] formerly of the

[1] Brown, Ruth A.: "A Study of High-School and College Records of Freshmen Entering Twenty Colleges and Universities in Michigan in the Fall of 1928." *Journal of Michigan Schoolmasters Club*, Vol. LXV, 1930, pp. 315–330.

Office of Educational Investigations of the University of Michigan. Page 325 of this report states that there is a correlation of .494 between average first semester college grades and grades in high school for the four thousand nine hundred fifty-one freshmen entering the twenty Michigan colleges. The correlation sheet was taken from the files and the subcorrelation coefficients computed. These are presented in Table VI. (A group of the 1928 University of Michigan freshmen is included in this study.)

It thus appears that the same tendency is evident in other Michigan institutions. For purposes of comparison there is presented in Table VII a list of the subcorrelations given above.

SUMMARY AND CONCLUSIONS

1. The subcorrelation technique is applicable to the problem of predicting college scholastic success from high-school record.

2. In Michigan the subgroup composed of the *A* and *B* high-school students has better predictive power than do other subgroups. The *C* subgroup has about average predictive power while the group of those below *C* has a low predictive power, since the line of regression has a large slope. This is presumably due to the fact that the admissions officers have admitted students from this group because other evidence outweighed the scholastic records.

3. It is probable that the pattern described (or variations of it) is quite universal. In any case the subcorrelation method can be used in discovering the pattern and providing comparable measures in cases where the distribution is non-linear, non-homoschedastic, or both.

COMPARATIVE SOCIAL ATTITUDES[1]

MAURICE M. SMITH

University of California

While the promotion and development of desirable social attitudes is widely accepted as a major objective of social-studies instruction in the schools, little effort has been made to determine the nature of these attitudes or to measure the amount of change resulting from such instruction. For the most part we teach facts and assume a concomitant development of attitudes.[2] But an attitude is in the nature of an evaluation, a point of view. There is a logical basis for questioning the inevitable concomitant origin of attitudes and certain evidence[3] has been reported which challenges this assumption.

A study of the social attitudes of high-school seniors was undertaken in California to throw light on these general problems, but more specifically to discover the nature of the social attitudes of graduating high-school seniors, and to compare their attitudes with those of high-school teachers, parents of high-school seniors, and professional social scientists in the universities and colleges of California.

Using approved techniques, there was constructed for the above purposes a social-science opinionnaire designed to sample significant aspects of five major social institutions, *viz.*, the family, the school, the church, industry, and the State. Two questions were paramount; first, what are the more significant issues related to these institutions, and, second, upon which of these should high-school seniors have informed opinions. The principal sources used for a selection of basic issues were the studies of Hockett,[4] Lee,[5] Billings,[6] and *Recent Social*

[1] The study here abstracted is reported in full in the writer's doctoral dissertation, *The Attitudes and Opinions of California High-School Seniors Toward Current Social Issues*, written under the direction of Professor Noel Keys at the University of California, 1935.

[2] Kelley, T. L.: *Scientific Method.* New York: Macmillan, 1932, p. 105ff.

[3] Longstreet, R. J.: "An Experiment with the Thurstone Attitude Scales." *School Review*, Vol. XLIII No. 3, March, 1935; pp. 202–210.

[4] Hockett, John A.: *A Determination of the Major Social Problems of American Life.* Teachers' College Contributions to Education. No. 281 New York: Bureau of Publications, Columbia University, 1927.

[5] Lee, Baldwin: *Issues on the Social Studies.* Teachers College Contributions to Education. New York: Bureau of Publications, Columbia University, 1928.

[6] Billings, Neal: *Determination of Generalizations Basic to the Social Studies Curriculum.* Baltimore: Warwick and York, 1928.

Trends.[1] Some three thousand issues were selected from these sources and classified under the headings of the five social institutions. Statements were written expressing opinions toward six hundred of these issues. This list was reduced to three hundred by eliminating overlapping issues.

The second question raised above was answered by competent judges.[2] An additional one hundred of the statements were eliminated through these judgments. Preliminary administrations of the remaining two hundred items disclosed eighty additional items which were unsatisfactory because of certain ambiguities of phrasing or standard deviations so large that group means would have proved of little value. In its final form the opinionnaire was composed of one hundred twenty significant controversial issues sampling aspects of our social institutions. Provision was made for recording individual opinion toward these statements as phrased on a five-point scale of agreement-disagreement.

Under standard conditions the opinionnaires were administered to the four populations under consideration in the spring of 1935. The first included eleven hundred seventy-six high-school seniors representing four urban, four town, and eight rural high-school districts of California. Two hundred eighty-eight parents of these seniors, one hundred ninety-two of their teachers, and eighty-three university professors from the academic departments of economics, political science, history, sociology, education, and psychology in leading California institutions constituted the other groups.

An item analysis was made of the opinions thus secured. The pupil population was sub-classified according to sex, and into rural, town, and urban groups. Using five categories of opinion toward each item, frequency distributions were made for each item for all groups. The following deviation values were assigned[3] to the opinion categories:

Strongly agree (with the statement as phrased) = +2
Agree (with reservations) = +1

---

[1] *White House Conference Report on Recent Social Trends.* New York: Century, 1934.

[2] The judgment of the California State Advisory Committee on Secondary Education was secured to afford an objective selection of statements and to insure that these statements should come within the training of the average high-school senior.

[3] Likert, Rensis: "A Technique for the Measurement of Attitudes." *Archives of Psychology*, No. 140, 1932.

Doubtful (equal uncertainty for or against)	= 0
Disagree (with reservations)	= −1
Strongly disagree	= −2

When the frequency distributions were weighted with these deviation values, mean and standard deviation scores were obtained for each group and each item. The preliminary treatment of these data revealed so few statistically reliable differences between the mean scores of the sub-groups of seniors that these were pooled and subsequently treated as one group.

The treatment of the data was largely restricted to comparisons between the mean scores of the groups toward the statements under consideration. Many highly significant differences were found. These have been reported elsewhere.[1]

A technique was needed which would express the total relationships among the four groups toward items relating to each of the five social institutions, and which would at the same time identify significant mean differences that are so apt to be obscured in any summation of test scores. The first of these requirements was met by adapting the usual correlation procedure to present data by pairing the mean scores of two groups toward the same items. As ordinarily used, however, the correlation technique has the weakness of any summation of test scores in that the identity of individual items is lost in the process. To overcome this inadequacy and to meet the second requirement stated above, a modification of the scatter diagram was made.[2] This consisted of writing the serial numbers of the items in the proper cells where the usual tally marks are placed.

ATTITUDES TOWARD INDUSTRY

The nature, form, and function which industry should take in the proximate future is a question that is engaging the attention of a thoughtful adult population. While no satisfactory solutions to present industrial problems have been forthcoming, yet the schools, through courses in civics or social studies, purport to adjust a younger generation to these adult realities.

As a result of all the forces that have been at work on the high-school senior, both in school and out, what attitudes and opinions has he formed toward contemporary industrial problems? Are these opinions

---

[1] Smith, Maurice M.: *Op. cit.*, p. 178ff.

[2] The author is indebted to Professor Robert C. Tryon, of the University of California, for suggesting this use of the scatter diagram.

similar to those of his parents, his teachers, or of professional students of such questions? Is there evidence that the school has been instrumental in formulating pupil opinions? Is there evidence, as some claim, that leftist teachers are creating a new social order by establishing extreme positions in the minds of our youth?

A tentative answer to such questions was attempted with thirty-six statements revealing group opinions toward the nature, form, and function of industry; toward the nature of desirable relationships between government and business, the private ownership of wealth, competition, business ethics, individualistic opportunities, working conditions, labor unions, and the like which were subsumed under the heading of industry. Statements such as the following were included under industry:

Opinion statement	Group means			
	Students	Parents	Teachers	Professors
Newspapers exaggerate greatly when they state that 5 per cent of the people own 95 per cent of all the wealth in this country	.13	.12	−.41	−.48
The federal government should regulate advertising and prohibit the publication of false or misleading statements	.81	.96	.96	.91
Everyone who earns his own money has a right to spend it as he pleases	.54	.50	−.03	−.48

Coefficients of correlation were obtained between each pair of the groups by pairing their attitude scores toward each item. The total relationships thus obtained are presented in Table I. When it is recalled that these thirty-six opinion statements were chosen originally because of their controversial nature, the magnitude of the coefficients becomes even more than ordinarily significant.

On questions of industry present data indicate a highly consistent agreement between high-school seniors and their parents, with only a slightly greater disparity between high-school teachers and university professors. Group opinions toward individual items reveal that parents favor private ownership of business with less governmental regulation, an extension of individualistic opportunities, and that they express, on the whole, lower standards of business ethics than do university professors and teachers. There is little evidence that the more

TABLE I.—COEFFICIENTS OF CORRELATION BETWEEN THE MEAN SCORES OF SENIORS, PARENTS, TEACHERS, AND PROFESSORS TOWARD THIRTY-SIX OPINION STATEMENTS RELATING TO INDUSTRY

	Groups		
	Seniors	Parents	Teachers
Parents	.96 ± .01		
Teachers	.88 ± .03	.90 ± .02	
Professors	.87 ± .04	.80 ± .04	.93 ± .02

highly socialized attitudes of the professional students of industrial questions are being accepted by these high-school seniors. The total variance between the groups is indicative of basic philosophical differences which conceivably could be highly influential in determining behavior.

ATTITUDES TOWARD THE STATE

That the present economic crisis is widening the functions of government beyond anything heretofore experienced in America is unquestioned. The assumption of the responsibility for providing work relief, social security, industrial regulation, agricultural control, and other recovery acts are interpreted by many as the beginning of a new and more collectivistic era in which an ever-increasing burden of the responsibility for the public welfare should be shifted to governmental agencies.

Thirty-eight items of the opinionnaire were concerned with attitudes toward the State. Included were attitudes toward the responsibility of the State in achieving the general welfare, the responsibilities and duties of citizens, the continuance of present governmental forms and divisions, the soundness of present court practices, the administration of justice, law enforcement, the influence and regulation of public opinion-forming agencies, and the general function of government in the present economic crisis. These statements may be illustrated by those given in the table at the top of page 686.

Total relationships between the opinions of the four groups toward the basic issues of citizenship and government are indicated in the correlation coefficients in Table II. These data indicate a greater disparity among the groups than was evident in the field of industry, but there is a fairly high correspondence between pupils and their

Opinion statement	Group means			
	Students	Parents	Teachers	Professors
Recent developments in all forms of transportation and communication have made unnecessary many of our existing forms of local government	.12	.38	.51	.79
Public officials, whose offices require skilled or professional training, should be appointed rather than elected	−.09	.23	.64	.93
After good men have been elected to office the average citizen should tend to his own affairs and let officials manage public affairs	.23	.48	−.26	−.32

TABLE II.—COEFFICIENTS OF CORRELATION BETWEEN THE MEAN SCORES OF SENIORS, PARENTS, TEACHERS, AND PROFESSORS TOWARD THIRTY-EIGHT OPINION STATEMENTS RELATING TO STATE AND GOVERNMENT

	Groups		
	Seniors	Parents	Teachers
Parents	.85 ± .03		
Teachers	.78 ± .04	.78 ± .04	
Professors	.45 ± .08	.36 ± .09	.72 ± .05

parents. On the surface these total relationships would seem to lend support to the popular notion that social scientists are, on the whole, proponents of political and social doctrines opposed to or subversive to accepted principles of American government. This would be especially true if the attitudes and opinions of parents, as a group, were taken as the standard of reference.

An inspection of those specific items toward which differences of opinion between parents (and pupils) and professors (and teachers) are most marked, reveals that while all groups favor a wider assumption of responsibility by the federal government for the promotion and financing of measures making for the general welfare of the people, economic security, work, sick and old age benefits, and the like, parents (and pupils) are consistently unwilling to assume a proportionate share of the responsibilities and the abridgments of personal rights and liberties necessarily associated with a wider dissemination of public benefits by the federal government. Social scientists on the other hand (and,

to a lesser extent, high-school teachers, also) tend to express attitudes which indicate a more adequate understanding of the conditions requisite to the promotion of the general welfare by a federal authority. On the part of parents, as a group, "the promotion of the general welfare" would seem to be a stereotyped phrase largely dissociated from reality.

High-school seniors seem largely unaware of the influence which modern developments of transportation and communication are having upon local governmental units. There is likewise little indication that they are informed with reference to modern methods of handling crime, the treatment of criminals, or legal procedure and court practices favorable to the administration of justice. They agree with their parents, that the established order is good and should be maintained.

On the whole, the opinions of the groups toward issues relating to state and government reveal a discouragingly negligible influence of the classroom in promoting attitudes among high-school seniors which are harmonious with those of more informed persons.

ATTITUDES TOWARD THE CHURCH AND PERSONAL MORALITY

Popular opinion depicts the youngsters of today as radically different from their elders in their ethical and moral standards, in the stability of their beliefs, in their attitudes toward religion, but there is meager evidence to indicate just how different the opinions of the present generation of high-school pupils are from their parents and teachers.

In the present study thirty-four items of the opinionnaire related to the church, to ethics, and ideals of character and conduct. Included were such statements as:

Opinion statement	Group means			
	Students	Parents	Teachers	Professors
Churches and ministers should confine their activities to religious matters and leave political affairs alone	.55	.66	−.03	−.48
The best rule for good conduct is to follow the promptings of one's conscience	.36	.45	.03	−.08
A person who is forced to accept a salary lower than his real worth has a right to loaf on the job to make up for it	−.86	−.85	−1.12	−.99

Toward the issues raised in this section the mean opinions of the four groups were, on the whole, in striking agreement. The total correspondence is shown in Table III.

TABLE III.—COEFFICIENTS OF CORRELATION BETWEEN THE MEAN SCORES OF SENIORS, PARENTS, TEACHERS, AND PROFESSORS TOWARD THIRTY-FOUR OPINION STATEMENTS RELATING TO THE CHURCH, RELIGION, AND ETHICS

	Groups		
	Seniors	Parents	Teachers
Parents	.97 ± .01		
Teachers	.91 ± .03	.97 ± .01	
Professors	.78 ± .05	.81 ± .05	.95 ± .01

Such opinion differences as were found toward these questions suggest a higher idealism and a greater appreciation of ethical, moral, and social values on the part of the teaching profession than on the part of high-school seniors and their parents, as a group. The differences between high-school seniors and their parents are slight. Popular characterizations of high-school youths as a rootless, opinionless generation widely at variance with their elders, seem unfounded if present data are to be relied upon.

ATTITUDES TOWARD THE FAMILY

A diminution of the influence of the family, as well as marked changes in the assumption of responsibilities by young persons in their family relationships are widely presumed[1] to have paralleled recent observable changes in the economic function and form of the family.

To discover the consensus with respect to this assumption, thirty-two items of the opinionnaire were selected to sample attitudes toward the function of persons in their family relationships, toward problems of marriage and divorce, illigitimacy, birth control, sex instruction, eugenics, the place of woman in industry, problems of home management, health, child care, discipline, duties, standard of living, and other similar issues.

The following statements illustrate the type of questions raised in this section:

[1] Queen, S. A.: *Social Organization and Disorganization.* New York: Thomas Y. Crowell Company, 1935.

Opinion statement	Group means			
	Students	Parents	Teachers	Professors
High schools should give detailed sex instruction to prepare students more adequately for marriage	.65	.54	.26	.37
Each community should establish Mother's Health Clinics to furnish mothers with scientific birth control information and to distribute supplies at cost or according to one's ability to pay	.86	.66	.74	1.03
A man should not ask a girl to marry him until he is earning enough to give her all the comforts and little luxuries that are so necessary to a woman's happiness	.58	.42	−.28	−.62

Certain high-school districts in California (Sacramento, Berkeley, and Alameda) objected to this section of the opinionnaire on the grounds that for the schools, these were unmentionable issues. The action of these schools is particularly interesting in view of the high agreement discovered between the groups toward these items. These relationships are indicated in Table IV. It is noteworthy that these

TABLE IV.—COEFFICIENTS OF CORRELATION BETWEEN THE MEAN SCORES OF SENIORS, PARENTS, TEACHERS, AND PROFESSORS TOWARD THIRTY-TWO OPINION STATEMENTS RELATING TO THE FAMILY

	Groups		
	Seniors	Parents	Teachers
Parents	.99 ± .004		
Teachers	.97 ± .01	.98 ± .01	
Professors	.64 ± .07	.65 ± .07	.95 ± .01

coefficients are among the highest uncovered in the present study and indicate an astonishing unanimity of opinion between all groups save professors with seniors and with parents. The evolution of the family from an independent economic unit to the present functionaly decentralized institution is an observable phenomenon. Observable, also, are certain trends of social disintegration. The assumption, however, of a causal relationship between the two trends is not supported in the

present study. Present data would seem to indicate that there has not been a marked disintegration of family influences, nor has there been a striking change in family ideals, at least, within one generation.

The opinion differences between the groups point toward a greater tolerance, a broader and more sympathetic understanding, and a higher idealism on the part of the professional social scientist.

ATTITUDES TOWARD THE SCHOOL

Educational planning commissions are attempting to formulate school policies with regard to questions of indoctrination, freedom of speech in the class room, the responsibility of the schools in perpetuating the established order, the function of the school in a changing society, the responsibility of the school in preparing youths for parenthood, for citizenship, and for social living in a competitive world.

Opinion statement	Mean scores			
	Students	Parents	Teachers	Professors
Teachers should teach only proved truths and should avoid discussing controversial issues in the classroom	−.44	.21	−.75	−.86
Courses of study have changed as rapidly as have the social and economic conditions outside the school	.29	.28	−.60	−.89
The use of modern machines in industry has largely eliminated the need for highly trained men	.12	.22	.−12	−.39

These problems loom large on the school frontier today, and much is being written about them by prominent educators. Thus, for example, Kefauver[1] states: "The educational program in a democracy should be in harmony with the dominant democratic social values and aspirations," and again, "the school has an unequivocal mandate to vitalize the democratic social ideal and to develop an educational program in harmony with it." Broad generalizations such as these tend to be accepted without question, but when specific issues are raised which apply the generalization, unanimity of opinion is not so easily obtained. The opinions of students, parents, teachers, and social scientists toward

[1] Kefauver, Grayson N.: "An Educational Platform for 1936." *School and Society*, Vol. XLIV, No. 1132, Sept. 5, 1936, pp. 289–295.

the functions and responsibilities of the school were sampled with thirty items of the opinionnaire. Included were such items as shown in the table on page 390.

The total relationships between the mean opinions of the groups are shown in Table V. It is evident that there is considerable disparity

TABLE V.—COEFFICIENTS OF CORRELATION BETWEEN THE MEAN SCORES OF SENIORS, PARENTS, TEACHERS, AND PROFESSORS TOWARD THIRTY OPINION STATEMENTS RELATING TO THE SCHOOL

	Groups		
	Students	Parents	Teachers
Parents	.81 ± .04		
Teachers	.81 ± .04	.87 ± .03	
Professors	.64 ± .07	.55 ± .09	.82 ± .04

between the opinions of parents and seniors on the one hand and the more informed social scientists on the other. Parents (and students) definitely favor the maintenance of the established order, with the implication that that which is, represents the true American Ideal. They favor rigorous restriction on freedom of speech and instruction, and are unwilling to have their children exposed to new social and political philosophies, favor biased and edited textbooks, and are of the opinion that high schools are developing in their students attitudes of critical open-mindedness toward social, political, and economic problems. In general there seems a well-defined attitude on the part of parents that there shall be local determination of what social ideals and values shall be inculcated by the school. The "dominant social values and aspirations" of the democracy are those which they themselves hold and approve.

TOTAL RELATIONSHIPS

Pooling all the items of the opinionnaire into one series, correlation coefficients were computed to indicate the total relationships between the groups. These coefficients are reported in Table VI. Teachers are seen to occupy a curious middle-ground between parents and professors. For the correlation coefficient of .88 there is an alienation coefficient of .48.[1] Reference to the items[2] toward which differences of

[1] $\sqrt{1 - r^2}$

[2] Space limitations prohibit the publication of the entire list of items in the present article.

TABLE VI.—COEFFICIENTS OF CORRELATION BETWEEN THE MEAN SCORES OF SENIORS, PARENTS, TEACHERS, AND PROFESSORS TOWARD THE TOTAL OPINIONNAIRE

	Groups		
	Seniors	Parents	Teachers
Parents	.92 ± .01		
Teachers	.85 ± .02	.88 ± .01	
Professors	.66 ± .04	.65 ± .04	.88 ± .01

opinion occur, indicate that when teachers disagree with parents they tend to agree with professors, and when they agree with professors they tend to disagree with parents. The greatest alienation, .75, occurs between parents and university professors and the least, .39, between parents and seniors.

THE RELATIVE STRENGTH OF GROUP OPINIONS

It is popularly believed that persons with little information basic to social and economic problems have strong prejudices regarding alternative solutions, and that professional students of these questions, being informed on both sides, tend to have weak or uncertain attitudes toward such propositions as those used in the present study. Nelson[1] has shown that the standard deviation of a series of mean scores may be used to indicate the relative strength of the opinions thus represented. A large standard deviation would thus indicate a tendency for the group to show a relatively high agreement in the direction of either positive or negative reactions toward the several opinion statements. The standard deviations for the group scores on the total opinionnaire were found to be:

SD students = .49
SD parents = .53
SD teachers = .63
SD professors = .66

It is noteworthy that, contrary to popular belief, university professors express most emphatic opinions and that, toward the issues presented, they show the closest agreement among themselves.

[1] Nelson, Ernest: *A Study of Conformity of Opinion in Certain College Groups.* Unpublished Master's Thesis, Lange Library, University of California, 1933, under the direction of Prof. Noel Keys, cf. p. 39ff.

CAUSAL RELATIONSHIPS

The assumption is widely made that adults exert a strong influence over high-school students. The development of desirable attitudes is stated as an objective of social-studies instruction in the schools, the presumption being that this goal is attainable. If a further assumption is made that this influence is in one direction only,[1] *i.e.*, from adults to the pupils, partial correlation coefficients derived from present data might be interpreted as indicating the dependence of pupil attitudes upon parent or teacher or professor attitudes when the influence of the others had been ruled out. Under this assumption the relationships thus revealed were found to be as follows:

The dependence of seniors on parents,	$r_{12.34} = .61$
The dependence of seniors on teachers,	$r_{13.24} = .12$
The dependence of seniors on professors,	$r_{14.23} = .01$
The multiple correlation coefficient,	$R_{1(234)} = .92$

SUMMARY

There is little evidence in the present study to indicate that social-studies instruction in our schools is succeeding in developing in our California high-school seniors attitudes and intelligent opinions toward issues basic to citizenship and social betterment beyond those which the seniors share with their parents as a group.

The opinions of no one group reflect an interpretation of social issues which would be accepted by other groups as an expression of the dominant social values and aspirations of the democracy. In view of the strength of these conflicting opinions, it would seem that any application of a restricting "frame of reference" to specific issues would inevitably compel the schools to attempt an indoctrination of attitudes and values contrary to those held by the parents of our high-school children.

The implication seems clear that if the schools are to promote and develop social attitudes in harmony with those held by persons best informed on social questions, cognizance must be taken of parental attitudes. There is strongly indicated a need for the development of coördinated classes for parents with their children to study and to discuss public affairs under the guidance of the schools in order that there may eventuate intelligent opinions and unbiased attitudes toward significant social problems.

[1] The judgment of the author is that causal influences are probably reciprocal, but that the influence of adults over the boys and girls undoubtedly predominates.

SEX DIFFERENCES IN PERFORMANCE ON THE AMERICAN COUNCIL PSYCHOLOGICAL EXAMINATION

T. M. LIVESAY

University of Hawaii

This paper is a continuation of a study reported in an earlier number of THE JOURNAL OF EDUCATIONAL PSYCHOLOGY.[1] The same data are here analyzed to show sex differences—(1) as a whole, and (2) as between the sexes according to race—and the reliability of such differences.

TABLE I.—MEANS, SIGMAS OF THE DISTRIBUTIONS, AND SIGMAS OF THE MEANS OF THE MALES (FIVE HUNDRED FIVE) AND FEMALES (THREE HUNDRED SEVENTY-TWO) FOR THE FIVE SUB-TEST SCORES AND TOTAL SCORES ON THE AMERICAN COUNCIL PSYCHOLOGICAL EXAMINATION

Test	Measure	Male	Female
Completion	Mean	27.15	26.35
	σ(dis.)	12.435	11.605
	σ(mean)	.553	.602
Artificial language	Mean	33.78	39.50
	σ(dis.)	15.180	18.000
	σ(mean)	.676	.933
Analogies	Mean	26.60	24.50
	σ(dis.)	14.420	13.915
	σ(mean)	.642	.722
Arithmetic	Mean	27.92	17.99
	σ(dis.)	12.935	11.935
	σ(mean)	.576	.619
Opposites	Mean	33.56	35.05
	σ(dis.)	16.150	16.630
	σ(mean)	.719	.862
Total score	Mean	148.72	143.89
	σ(dis.)	48.345	50.700
	σ(mean)	2.151	2.629

[1] Livesay, T. M.: "Racial Comparisons in Performance on the American Council Psychological Examination." *Journal of Educational Psychology*, November, 1936.

Table I gives the means, sigmas of the distributions, and sigmas of the means for the males as a whole (five hundred five) and females as a whole (three hundred seventy-two) for the sub-test scores and total scores—basic data for the computations of the chances of true differences as given in Table II. It will be noted that the differences are in favor of the males (Table II) in four of the comparisons, and of the females in two; that for the Artificial Language and Arithmetic tests there is complete reliability of the differences; and that the difference in the case of the Analogies test shows almost complete reliability as the chances are ninety-nine in one hundred. As between high-school graduates it would seem that on the whole men do somewhat better on the American Council Psychological Examination than do women.

TABLE II.—SEX COMPARISONS (TOTALS) IN TERMS OF THE CHANCES OF A TRUE DIFFERENCE

Tests	Difference between means D	σ(diff.)	Critical ratio $\frac{D}{\sigma(\text{diff.})}$	Chances in one hundred that the true difference is greater than zero[1]	Difference in favor of
Completion	.80	.818	.98	84	Male
Artificial language	5.72	1.152	4.97	100	Female
Analogies	2.10	.966	2.17	99	Male
Arithmetic	9.93	.845	11.75	100	Male
Opposites	1.49	1.122	1.33	91	Female
Total score	4.83	3.396	1.42	92	Male

[1] Garrett, Henry E.: *Statistics in Psychology and Education.* Table XIV, p. 134.

In Table III the basic data are presented—means, sigmas of the distributions, and sigmas of the means of the males and females of the four racial groups for the sub-test and total scores—for the computations in Table IV of the chances of true differences.

An analysis of Table IV indicates differences in favor of:

(1) Caucasian males over Caucasian females in four tests out of six, with complete reliability of the difference in favor of the females in the Artificial Language test and complete reliability of the difference in favor of the males in the Arithmetic test;

(2) Caucasian males over Chinese females in all but Artificial Language, with complete reliability of the differences in four cases;

(3) Caucasian males over Japanese females in all but Artificial Language, with complete reliability in every case;

(4) Caucasian males over Part-Hawaiian females in all tests with complete reliability in all but Artificial Language;

(5) Caucasian females over Chinese males in all tests but Arithmetic, with complete reliability in three cases;

(6) Chinese males over Chinese females in all but Opposites, with complete reliability only in the case of Arithmetic;

(7) Chinese males over Japanese females in all but Artificial Language, with complete reliability in two cases;

(8) Chinese males over Part-Hawaiian females in all tests with complete reliability in three cases;

(9) Caucasian females over Japanese males in all but Arithmetic; with complete reliability in all cases except Analogies;

TABLE III.—NUMBERS, MEANS, SIGMAS OF THE DISTRIBUTIONS, AND SIGMAS OF THE MEANS OF THE MALES AND FEMALES OF THE FOUR RACIAL GROUPS FOR THE FIVE SUB-TEST SCORES AND THE TOTAL SCORES ON THE AMERICAN COUNCIL PSYCHOLOGICAL EXAMINATION

Test	Measure	Caucasian		Chinese		Japanese		Part-Hawaiian	
		Male, 144	Female, 121	Male, 94	Female, 84	Male, 194	Female, 98	Male, 51	Female, 46
Completion	Mean	34.12	31.25	25.25	25.22	22.74	21.90	27.30	23.95
	σ(dis.)	12.660	12.425	12.550	10.510	10.385	9.975	10.495	9.695
	σ(mean)	1.060	1.129	1.295	1.147	.746	1.008	1.469	1.429
Artificial language	Mean	31.75	40.64	34.98	32.18	34.88	39.40	30.63	31.13
	σ(dis.)	14.975	14.965	16.445	18.165	14.665	18.365	12.050	14.900
	σ(mean)	1.248	1.360	1.696	1.982	1.053	1.855	1.687	2.197
Analogies	Mean	29.81	29.27	27.25	21.94	24.86	23.89	22.79	19.28
	σ(dis.)	14.565	15.310	14.545	14.235	14.230	13.390	13.075	11.215
	σ(mean)	1.214	1.392	1.500	1.553	1.022	1.353	1.831	1.654
Arithmetic	Mean	28.60	20.30	31.30	18.09	27.85	16.85	20.53	15.04
	σ(dis.)	12.620	15.525	14.470	13.545	11.960	10.845	12.535	8.755
	σ(mean)	1.052	1.139	1.493	1.478	.859	1.095	1.755	1.291
Opposites	Mean	39.05	40.47	31.90	33.19	31.38	31.34	28.96	31.57
	σ(dis.)	17.075	17.640	14.415	15.990	15.165	14.455	15.690	16.010
	σ(mean)	1.423	1.604	1.487	1.745	1.089	1.460	2.197	2.361
Total score	Mean	162.31	161.71	151.40	141.82	141.85	134.05	131.41	120.15
	σ(dis.)	52.855	53.685	48.860	55.575	44.475	41.110	37.365	46.620
	σ(mean)	4.405	4.880	3.508	6.064	3.193	4.153	5.233	6.874

TABLE IV.—SEX COMPARISONS OF THE FOUR RACIAL GROUPS IN TERMS OF THE CHANCES OF A TRUE DIFFERENCE

Race combinations	Tests	Difference between means D	σ(diff.)	Critical ratio $\frac{D}{\sigma(\text{diff.})}$	Chances in one hundred that the true difference is greater than zero[1]	Difference in favor of
Caucasian male Caucasian female	Completion	2.87	1.549	1.85	96	Caucasian male
	Artificial language	8.89	1.846	4.82	100	Caucasian female
	Analogies	.54	1.847	.29	62	Caucasian male
	Arithmetic	8.30	1.550	5.35	100	Caucasian male
	Opposites	1.42	2.144	.66	74	Caucasian female
	Total score	.60	6.574	.09	54	Caucasian male
Caucasian male Chinese female	Completion	8.90	1.562	5.69	100	Caucasian male
	Artificial language	.43	2.341	.18	57	Chinese female
	Analogies	7.87	1.971	3.99	100	Caucasian male
	Arithmetic	10.51	1.814	5.79	100	Caucasian male
	Opposites	5.86	2.252	2.60	99	Caucasian male
	Total score	20.49	7.495	2.73	100	Caucasian male
Caucasian male Japanese female	Completion	12.22	1.463	8.35	100	Caucasian male
	Artificial language	7.65	2.236	3.42	100	Japanese female
	Analogies	5.92	1.818	3.26	100	Caucasian male
	Arithmetic	11.75	1.518	7.74	100	Caucasian male
	Opposites	7.71	2.029	3.78	100	Caucasian male
	Total score	28.26	6.054	4.67	100	Caucasian male
Caucasian male Part-Hawaiian female	Completion	10.17	1.779	5.72	100	Caucasian male
	Artificial language	.62	2.527	.25	60	Caucasian male
	Analogies	10.53	2.052	5.13	100	Caucasian male
	Arithmetic	13.56	1.665	8.14	100	Caucasian male
	Opposites	7.48	2.757	2.71	100	Caucasian male
	Total score	42.16	8.164	5.16	100	Caucasian male
Chinese male Caucasian female	Completion	6.00	1.718	3.49	100	Caucasian female
	Artificial language	5.66	2.174	2.60	99	Caucasian female
	Analogies	2.02	2.046	.99	84	Caucasian female
	Arithmetic	11.00	1.878	5.86	100	Chinese male
	Opposites	8.57	2.187	3.92	100	Caucasian female
	Total score	10.31	6.010	1.72	96	Caucasian female
Chinese male Chinese female	Completion	.03	1.730	.02	51	Chinese male
	Artificial language	2.80	2.609	1.07	85	Chinese male
	Analogies	5.31	2.159	2.46	99	Chinese male
	Arithmetic	13.21	2.101	6.29	100	Chinese male
	Opposites	1.29	2.293	.56	71	Chinese female
	Total score	9.58	7.006	1.37	91	Chinese male
Chinese male Japanese female	Completion	3.35	1.642	2.04	98	Chinese male
	Artificial language	4.42	2.513	1.76	96	Japanese female
	Analogies	3.36	2.020	1.66	95	Chinese male
	Arithmetic	14.45	1.852	7.80	100	Chinese male
	Opposites	.56	2.084	.27	61	Chinese male
	Total score	17.35	5.436	3.19	100	Chinese male
Chinese male Part-Hawaiian female	Completion	1.30	1.928	.67	75	Chinese male
	Artificial language	3.85	2.775	1.39	92	Chinese male
	Analogies	7.97	2.233	3.57	100	Chinese male
	Arithmetic	16.26	1.974	8.24	100	Chinese male
	Opposites	.33	2.790	.12	55	Chinese male
	Total score	31.25	7.717	4.05	100	Chinese male
Japanese male Caucasian female	Completion	8.51	1.353	6.29	100	Caucasian female
	Artificial language	5.76	1.720	3.35	100	Caucasian female
	Analogies	4.41	1.727	2.55	99	Caucasian female
	Arithmetic	7.55	1.427	5.29	100	Japanese male
	Opposites	9.09	1.939	4.69	100	Caucasian female
	Total score	19.86	5.832	3.41	100	Caucasian female

[1] Garrett, Henry E.: *Statistics in Psychology and Education*, Table XIV, p. 134.

TABLE IV.—*Continued*

Race combinations	Tests	Difference between means D	σ(diff.)	Critical ratio $\frac{D}{\sigma(\text{diff.})}$	Chances in one hundred that the true difference is greater than zero[1]	Difference in favor of
Japanese male Chinese female	Completion	2.48	1.368	1.81	96	Chinese female
	Artificial language	2.70	2.244	1.20	88	Japanese male
	Analogies	2.92	1.859	1.57	94	Japanese male
	Arithmetic	9.76	1.709	5.71	100	Japanese male
	Opposites	1.81	2.057	.88	81	Chinese female
	Total score	.03	6.849	.004	50	Japanese male
Japanese male Japanese female	Completion	.84	1.254	.67	75	Japanese male
	Artificial language	4.52	2.127	2.13	98	Japanese female
	Analogies	.97	1.697	.57	72	Japanese male
	Arithmetic	11.00	1.391	7.91	100	Japanese male
	Opposites	.04	1.821	.02	51	Japanese male
	Total score	7.80	5.238	1.49	93	Japanese male
Japanese male Part-Hawaiian female	Completion	1.21	1.612	.75	77	Part-Hawaiian female
	Artificial language	3.75	2.436	1.54	93	Japanese male
	Analogies	5.58	1.944	2.87	100	Japanese male
	Arithmetic	12.81	1.551	8.26	100	Japanese male
	Opposites	.19	2.600	.07	53	Part-Hawaiian female
	Total score	21.70	7.579	2.87	100	Japanese male
Part-Hawaiian male Caucasian female	Completion	3.95	1.853	2.13	98	Caucasian female
	Artificial language	10.01	2.167	4.62	100	Caucasian female
	Analogies	6.48	2.300	2.82	100	Caucasian female
	Arithmetic	.23	2.092	.11	51	Part-Hawaiian male
	Opposites	11.51	2.720	4.23	100	Caucasian female
	Total score	30.30	7.155	4.23	100	Caucasian female
Part-Hawaiian male Chinese female	Completion	2.08	1.864	1.12	86	Part-Hawaiian male
	Artificial language	1.55	2.603	.60	73	Chinese female
	Analogies	.85	2.401	.35	64	Part-Hawaiian male
	Arithmetic	2.44	2.294	1.06	85	Part-Hawaiian male
	Opposites	4.23	2.806	1.51	93	Chinese female
	Total score	10.41	8.010	1.30	90	Chinese female
Part-Hawaiian male Japanese female	Completion	5.40	1.782	3.03	100	Part-Hawaiian male
	Artificial language	8.77	2.507	3.50	100	Japanese female
	Analogies	1.10	2.277	.48	68	Japanese female
	Arithmetic	3.68	2.069	1.78	96	Part-Hawaiian male
	Opposites	2.38	2.638	.90	82	Japanese female
	Total score	2.64	6.681	.40	65	Japanese female
Part-Hawaiian male Part-Hawaiian female	Completion	3.35	2.049	1.63	95	Part-Hawaiian male
	Artificial language	.50	2.770	.18	57	Part-Hawaiian female
	Analogies	3.51	2.467	1.42	92	Part-Hawaiian male
	Arithmetic	5.49	2.179	2.52	99	Part-Hawaiian male
	Opposites	2.61	3.225	.81	79	Part-Hawaiian female
	Total score	11.26	8.639	1.30	90	Part-Hawaiian male

[1] GARRET, HENRY E.: *Statistics in Psychology and Education*, Table XIV, p 134.

(10) Japanese males over Chinese females in four tests out of six, with complete reliability only in the case of Arithmetic;

(11) Japanese males over Japanese females in all but Artificial Language, with complete reliability only in the case of Arithmetic;

(12) Japanese males over Part-Hawaiian females in four out of six tests with complete reliability in three cases;

(13) Caucasian females over Part-Hawaiian males in all but Arithmetic, with complete reliability in four cases;

(14) Part-Hawaiian males over Chinese females in Completion, Analogies, and Arithmetic and Chinese females over Part-Hawaiian males in Artificial Language, Opposites and Total Score, with no case of complete reliability;

(15) Japanese females over Part-Hawaiian males in four out of six tests, with complete reliability in two cases;

(16) Part-Hawaiian males over Part-Hawaiian females in four out of six tests with no case of complete reliability.

Table V gives the comparisons of the males of the four racial groups in terms of the chances of a true difference. The differences are in favor of the:

(1) Caucasian over the Chinese in four out of six tests with complete reliability in two cases;

(2) Caucasian over the Japanese in all but Artificial Language, with complete reliability in all but two cases;

(3) Caucasian over the Part-Hawaiian in all tests with complete reliability in all but Artificial Language;

(4) Chinese over Japanese in all tests with no case of complete reliability;

(5) Chinese over Part-Hawaiian in all but Completion with complete reliability in two cases only;

(6) Japanese over Part-Hawaiian in all but Completion, with complete reliability in two cases only.

In Table VI the females of the four racial groups are compared. The differences are in favor of the:

(1) Caucasian over the Chinese in all tests with complete reliability in all but two cases;

(2) Caucasian over the Japanese in all tests with complete reliability in all but two cases;

(3) Caucasian over the Part-Hawaiian in all tests with complete reliability in every case;

TABLE V.—COMPARISONS OF THE MALES OF THE FOUR RACIAL GROUPS IN TERMS OF THE CHANCES OF A TRUE DIFFERENCE

Race combinations	Tests	Difference between means D	σ(diff.)	Critical ratio $\frac{D}{\sigma(\text{diff.})}$	Chances in one hundred that the true difference is greater than zero[1]	Difference in favor of
Caucasian Chinese	Completion	8.87	1.673	5.30	1.00	Caucasian
	Artificial language	3.23	2.106	1.53	93	Chinese
	Analogies	2.56	1.929	1.33	91	Caucasian
	Arithmetic	2.70	1.826	1.48	93	Chinese
	Opposites	7.15	2.058	3.47	100	Caucasian
	Total score	10.91	5.631	1.94	97	Caucasian
Caucasian Japanese	Completion	11.38	1.292	8.81	100	Caucasian
	Artificial language	3.13	1.633	1.92	97	Japanese
	Analogies	4.95	1.587	3.11	100	Caucasian
	Arithmetic	.75	1.358	.55	71	Caucasian
	Opposites	7.67	1.792	4.28	100	Caucasian
	Total score	20.46	5.440	3.76	100	Caucasian
Caucasian Part-Hawaiian	Completion	6.82	1.811	3.77	100	Caucasian
	Artificial language	1.12	2.098	.53	70	Caucasian
	Analogies	7.02	2.197	3.19	100	Caucasian
	Arithmetic	8.07	2.046	3.94	100	Caucasian
	Opposites	10.09	2.618	3.85	100	Caucasian
	Total score	30.90	6.841	4.52	100	Caucasian
Chinese Japanese	Completion	2.51	1.494	1.68	96	Chinese
	Artificial language	.10	1.996	.05	52	Chinese
	Analogies	2.39	1.815	1.32	90	Chinese
	Arithmetic	3.45	1.722	2.01	98	Chinese
	Opposites	.52	1.843	.28	61	Chinese
	Total score	9.55	4.743	2.01	98	Chinese
Chinese Part-Hawaiian	Completion	2.05	1.959	1.05	85	Part-Hawaiian
	Artificial language	4.35	2.392	1.82	96	Chinese
	Analogies	4.46	2.367	1.88	97	Chinese
	Arithmetic	10.77	2.304	4.67	100	Chinese
	Opposites	2.94	2.653	1.11	86	Chinese
	Total score	19.99	6.300	3.17	100	Chinese
Japanese Part-Hawaiian	Completion	4.56	1.647	2.77	100	Part-Hawaiian
	Artificial language	4.25	1.989	2.14	98	Japanese
	Analogies	2.07	2.097	.99	84	Japanese
	Arithmetic	7.32	1.953	3.75	100	Japanese
	Opposites	2.42	2.452	.99	84	Japanese
	Total score	10.44	6.131	1.70	96	Japanese

[1] Garrett, Henry E.: *Statistics in Psychology and Education.* Table XIV, p. 134.

TABLE VI.—COMPARISONS OF THE FEMALES OF THE FOUR RACIAL GROUPS IN TERMS OF THE CHANCES OF A TRUE DIFFERENCE

Race combinations	Tests	Difference between means D	σ(diff.)	Critical ratio $\frac{D}{\sigma(\text{diff.})}$	Chances in one-hundred that the true difference is greater than zero[1]	Difference in favor of
Caucasian Chinese	Completion	6.03	1.610	3.75	100	Caucasian
	Artificial language	8.46	2.404	3.52	100	Caucasian
	Analogies	7.33	2.076	3.53	100	Caucasian
	Arithmetic	2.21	1.866	1.18	88	Caucasian
	Opposites	7.28	2.370	3.07	100	Caucasian
	Total score	19.89	7.784	2.56	99	Caucasian
Caucasian Japanese	Completion	9.35	1.514	6.17	100	Caucasian
	Artificial language	1.24	2.300	.54	71	Caucasian
	Analogies	5.38	1.941	2.77	100	Caucasian
	Arithmetic	3.45	1.580	2.18	98	Caucasian
	Opposites	9.13	2.169	4.21	100	Caucasian
	Total score	27.66	6.408	4.32	100	Caucasian
Caucasian Part-Hawaiian	Completion	7.30	1.821	4.01	100	Caucasian
	Artificial language	9.51	2.583	3.68	100	Caucasian
	Analogies	9.99	2.161	4.62	100	Caucasian
	Arithmetic	5.26	1.722	3.05	100	Caucasian
	Opposites	8.90	2.854	3.12	100	Caucasian
	Total score	41.56	8.431	4.93	100	Caucasian
Chinese Japanese	Completion	3.32	1.527	2.17	99	Chinese
	Artificial language	7.22	2.715	2.66	99	Japanese
	Analogies	1.95	2.060	.95	83	Japanese
	Arithmetic	1.24	1.839	.67	74	Chinese
	Opposites	1.85	2.275	.81	79	Chinese
	Total score	7.77	7.350	1.06	85	Chinese
Chinese Part-Hawaiian	Completion	1.27	1.832	.69	76	Chinese
	Artificial language	1.05	2.959	.35	64	Chinese
	Analogies	2.66	2.269	1.17	87	Chinese
	Arithmetic	3.05	1.962	1.55	93	Chinese
	Opposites	1.62	2.936	.55	71	Chinese
	Total score	21.67	9.166	2.36	99	Chinese
Japanese Part-Hawaiian	Completion	2.05	1.749	1.17	87	Part-Hawaiian
	Artificial language	8.27	2.875	2.88	100	Japanese
	Analogies	4.61	2.137	2.16	98	Japanese
	Arithmetic	1.81	1.693	1.07	85	Japanese
	Opposites	0.23	2.776	.08	53	Part-Hawaiian
	Total score	13.90	8.031	1.73	96	Japanese

[1] Garrett, Henry E.: *Statistics in Psychology and Education.* Table XIV, p. 134.

(4) Chinese over the Japanese in all but two tests with no case of complete reliability;

(5) Chinese over the Part-Hawaiian in all tests with no case of complete reliability;

(6) Japanese over Part-Hawaiian in four of the six tests with complete reliability only in the case of Artificial Language.

In conclusion it should be pointed out, as in the previous article, that while these data indicate numerous sex differences of complete reliability, any significance which they may have as between races would apply only to these groups in the Territory of Hawaii and Continental United States. It is very doubtful if many of the racial groups in the United States (especially in the Territory of Hawaii) are to be considered as completely representative of their races as a whole because of the selective factors known to have been at work.

RESPONSE TO VISUAL *VS.* VISUAL-AUDITORY PRESENTATION IN A GO-TO-COLLEGE PROGRAM

FRANK R. ELLIOTT

Indiana University

The child gets its sense impressions of the world into which it is born largely through the ear. At about eight years of age, when it learns to read, the child's habits change, according to Kirkpatrick[1] and many other investigators, and the advantage swings away from auditory to visual mode. Further, it appears from various[2] studies that a combination of eye and ear stimulation becomes more effective upon attention and memory than does either mode alone.

While the above conclusions have been put to test often in laboratory experiments, they have been demonstrated far less frequently in the emotional frame of the life situation. How is the individual affected by the eye or ear appeal, or a combination of the two, when he is in his normal, everyday surroundings and has before him a practical problem to be solved, such as buying, management, decisions concerning a future career? To which mode of presentation does he react most effectively?

Some answer to the question seems to be indicated from a series of visual and visual-auditory types of presentation before ninety-two thousand six hundred eighty-one Indiana high-school seniors during the period, 1933–1935. Visual presentation was by printed literature; visual-auditory presentation was by literature combined with voice presentation. Contacts with the high-school seniors were under the author's direction for the purpose of informing the seniors as to facilities offered by their State university, Indiana University, for a higher education.

It is a common practice for universities and colleges, particularly the private institutions, to send literature and speakers to the high schools and other preparatory schools calling attention to courses, costs, and other facts of the respective institutions. Many high schools hold annual "Go-To-College" days and invite in college speakers. Tons of literature are mailed annually to the high-school seniors of the

---

[1] Kirkpatrick, E. A.: "Studies in Development and Learning." *Arch. Psychol.*, 1909, No. 12.

[2] Elliott, Frank R.: "Memory for Visual, Auditory and Visual-Auditory Material." *Arch. Psychol.*, 1936, No. 199, pp. 8–13.

nation. Generally speaking, it is a respected practice for representatives of institutions of higher learning and of high schools as well to do what they can to stimulate young men and women of intellect and determination to go forward to a college career. Ex-President Hoover was attracted to college in that way.

In the midst of the economic depression when jobs were scarce it was apparent that one of the most profitable uses the capable high-school graduate could make of his time was to try to improve himself through advanced study. Indiana University undertook to inform every Indiana high-school senior of its various courses, its costs, its opportunities for self-help, and other pertinent facts. A careful check of two types of response was made over the three-year period, 1933–1935: First, the inquiry response in the form of signed requests for departmental information, and, second, the enrollment response through registration at Indiana University the following September.

MATERIALS AND PROCEDURE

Visual materials included small, sixteen-page, descriptive, illustrated bulletins. Visual-auditory materials included this same literature plus the facts of the literature presented through the spoken word of faculty representatives. About forty faculty representatives, chosen in many cases because of their availability as much as because of their speaking ability, made the voice presentation. In view of the actual presence of the speaker, the visual-auditory mode in the present investigation represented a degree of social stimulation stronger than that which would be effected through such a visual-auditory medium as television or sound films.

It should be remembered that literature was furnished for students, whether or not a University representative was on hand. In each piece of literature was an inquiry card for the senior to use in requesting departmental information, and the number of these cards received was regarded as the measure of the inquiry response. Follow-up literature was mailed to those sending in the inquiry cards, whether those cards came from the visual or the visual-auditory group; follow-up procedure was the same for each group.

The visual group consisted of fifteen thousand five hundred ten seniors. The visual-auditory group covered seventy-seven thousand one hundred seventy-one seniors. It was impossible to send speakers to all the senior classes, and the effort was made to schedule them for small and large schools alike in a fairly even geographical distribu-

tion among approximately eight hundred Indiana high schools. Difficulties of talking to a senior class of several hundred students were such that the extremely large schools were sometimes omitted from the visual-auditory group. The difficulties of travel to remote sections were such that the extremely small schools were omitted at times from the visual-auditory group. An effort was made to rotate visits among the different schools of a county so that each school would be reached with visual-auditory presentation (literature and speaker) at least once in the three-year period, though this was not always possible in each county.

As in all field experimentation, it was impossible to control all conditions except the independent variable of mode of presentation. The numbers were so large, however, that it seems fairly reasonable to assume that such variables as influence of alumni were distributed equally throughout the visual and the visual-auditory groups to bring about what the author has termed an *Equilibrium of Variables.* If this assumption is valid, the independent variable of mode of presentation can be measured in its effects upon the ninety-two thousand six hundred eighty-one subjects.

RESULTS

The following tables show for each type of presentation the numbers reached, the numbers responding, the percentage responding, and the ratio of advantage held by one mode of presentation over the other. Results for each of the three years are shown, as well as results for the three years combined. Results for both the inquiry and the enrollment responses are indicated.

Table I shows that the percentage of inquiries following visual-auditory presentation is seven times as high for the three-year period as the percentage of inquiries following simple visual presentation. The advantage for visual-auditory presentation is marked for each of the three years.

Table II shows that the percentage of enrollments following visual-auditory presentation is twice as high as the percentage following simple visual presentation. The visual-auditory advantage is consistently pronounced for each of the three separate years.

These results bear out the author's[1] laboratory experiments and those of many other investigators showing the relative superiority of

[1] *Op. cit.*

TABLE I.—THE INQUIRY RESPONSE

Type of presentation	Numbers reached	Numbers responding	Percentage responding	Percentage advantage and ratio
Year 1933:				
Visual	8,430	353	4.18	Visual-auditory
Visual-auditory	21,847	5,031	23.03	5.3 to 1
Year 1934:				
Visual	4,149	316	7.61	Visual-auditory
Visual-auditory	27,102	9,989	36.86	4.8 to 1
Year 1935:				
Visual	2,931	14	.48	Visual-auditory
Visual-auditory	28,222	9,113	32.28	67.2 to 1
Three Years 1933–1935:				
Visual	15,510	683	4.44	Visual-auditory
Visual-auditory	77,171	24,133	31.27	7.0 to 1

TABLE II.—THE ENROLLMENT RESPONSE

Type of presentation	Numbers reached	Number responding	Percentage responding	Percentage advantage and ratio
Year 1933:				
Visual	8,430	144	1.70	Visual-auditory
Visual-auditory	21,847	941	4.30	2.5 to 1
Year 1934:				
Visual	4,149	97	2.34	Visual-auditory
Visual-auditory	27,102	1,076	3.97	1.7 to 1
Year 1935:				
Visual	2,931	89	3.04	
Visual-auditory	28,222	1,358	4.84	1.6 to 1
Three years 1933–1935:				
Visual	15,510	330	2.19	
Visual-auditory	77,171	3,375	4.37	2.0 to 1

combined visual and auditory stimulation as compared with simple visual mode. The sound of the human voice appears to have a social significance and appeal which is out of all proportion to its mere sensory stimulus value, and in our present study the presence of the speaker in person doubtless added force to the visual-auditory presentation.

The consistency of the results over three successive years, with nothing approaching a reversal of evidence favoring the visual-auditory

mode, seems to indicate the reliability of our findings. The enormous size of our test group, namely, ninety-two thousand six hundred eighty-one subjects, and the random distribution of the two modes of presentation among seniors of city and country schools, of poor and rich communities, of schools near to and distant from the University—all of this would tend to establish an equilibrium of variables and thus to rule out factors of influence other than that of mode of presentation.

Some evidence of the validity of our findings in the present field study is to be found in the fact that the field representative plan of building college enrollment has long been used successfully by the universities and colleges. Secondary-school officials have recognized the value of visual-auditory presentation through college field representatives, even to the extent of setting up "Go-To-College" programs and inviting speakers from universities and colleges to come to the high schools and present the facts of their respective institutions.

While common sense and experience have dictated the use of the visual-auditory mode or field representative plan of building college enrollment, the present study, so far as the author knows, offers the first objective evidence of the overwhelming advantage of the visual-auditory mode over the simple visual mode. In fact, with a small control group of three hundred seventy-one seniors the first year of the experiment, the author found sending literature alone very little better from the standpoint of gaining enrollment than doing nothing at all. The control group, which was not stimulated by the University in any way, contributed 1.62 per cent of its members to the University's enrollment, while the visual group contributed 1.7 per cent. This may be interpreted to mean that sending out vast quantities of printed matter, without any accompanying personal contact, is a great waste.

The recent national conference on education by radio emphasized the importance of the radio voice not so much from the standpoint of its value in formal instruction as from its importance as a *stimulus* to educational thought. So in the present experiment, the personalized, visual-auditory mode of presentation brought from two to seven times the response given to simple visual mode.

BOOK REVIEWS

Books on Personality

Winifred V. Richmond. *Personality, Its Development and Hygiene.* New York: Farrar & Rinehart, 1937.

Ross Stagner. *Psychology of Personality.* New York: McGraw-Hill Book Co., 1937.

James S. Plant. *Personality and the Cultural Pattern.* New York: The Commonwealth Fund, 1937.

The first of these recent books on personality is a presentation of modern approaches to the study of human nature by the psychologist of St. Elizabeth's Hospital. The personality is here presented in its development and hygiene. The book is based upon some of the material accumulated by the author in teaching various aspects of the subject throughout fifteen years to social workers, naval medical officers, nurses, medical students, teachers, parents, and in her practice with patients in a psychiatric hospital. The treatment is consequently colored by the psychiatric point of view. Dr. Richmond recognizes that scientific data on the human personality are as yet too scanty to make a comprehensive work on the subject possible. Nevertheless, whatever we do know about the human personality has come mainly from psychological investigation, so that a treatment of the subject must draw upon psychology. By personality Dr. Richmond means the physical and mental make-up that distinguish one person from another, and the body of her book is devoted to a discussion of the fundamental psychophysical traits: The body, intelligence, emotion, energy and orientation. The book also treats of the methods of studying the personality and the maladjustments and disorders of the personality.

Dr. Stagner's book is written from the standpoint of the psychologist interested primarily, if not exclusively, in the study of the normal, everyday personality, and is intended to meet the needs of those persons who complain that reading lists on the subject place too much stress on the pathological. The book deals in the main with personality traits and seeks to present a systematic set of concepts which will fill out the framework of trait psychology, measurement, procedures, etc. Personality is treated as a set of concrete, identifiable responses of a behaving organism, and hence stress is placed upon visceral processes and gestural activities. Another feature of the book is the contention of the author that plural statements have no real place in the psychology

of personality. The object of his study is a simple human being who reacts to and reflects the presence and characteristics of other human beings. A radical departure from the customary material found in books on this subject is a final chapter devoted to the problems of social reform. Psychology the author holds to be the basis of the social sciences, and hence it has something to contribute to a correct solution of social problems. It is in this spirit that he presents the deductions from personality to social values in that chapter.

The third of these books is by the Director of the Essex County Juvenile Clinic and is written with the conviction that "the usual psychiatric formulations for the problems of children are inadequate, and that in some way the forces of the cultural pattern in which we live are of dynamic importance to the personality." The author is, therefore, concerned throughout his book with a study of the impact of the environment upon inner motivations. The first three chapters are devoted to pointing out "what can be learned from an individual at the moment of his first and perhaps temporary break with his environment . . . " This is followed by a description of an effective set-up for tapping this source of information. The next four chapters give examples of the type of material that was gathered by this method. Nine chapters are then given over to tracing the implications of the author's hypothesis for the family, school, recreation, law and order, social work, medicine, church and industry. The final chapter brings the whole into focus in terms of education for life.

MAX SCHOEN.

Carnegie Institute of Technology.

HORATIO H. NEWMAN, FRANK N. FREEMAN and KARL J. HOLZINGER. *Twins; a Study of Heredity and Environment.* Chicago: Chicago University Press, 1937, pp. XVI + 369.

This long-awaited study scarcely lives up to our expectations. It is divided into two parts, the first a comparison of one hundred pairs of mono-zygotic and di-zygotic twins reared together; the second a comparison of twenty sets of mono-zygotic twins reared apart, nineteen of whom were tested by the authors. It is the second study that is the more interesting, since it is with identical twins reared apart that we should get a clear-cut issue as to amount of change that differing environments can effect. Take the case of the first pair, Alice and Olive, for instance. These were born in London, England. On the

death of the mother when the twins were eighteen months old, Olive was brought to Canada and Alice remained in London. The twins did not meet for seventeen years, but at the time of testing had lived together for a year in Canada. Since they were mono-zygotic they had identical heredities. What is the environmental effect of schooling in London compared with that of Oshawa? One has only to select suitable tests to determine the differences.

The experiment was carefully controlled. One had to be sure that the twins were mono-zygotic and the tests made to ensure that such was the case (p. 35) are beyond cavill. Secondly, the tests and measurements should be fair to both. That they were eminently fair to both can be seen from the tests and ratings used (see pages 29–31). Thirdly, the statistical comparisons of results should be above suspicion, and again the study meets the test.

It is when we come to the interpretation of the results that there is room for differences of opinion. Even the three authors are not altogether in agreement. Newman, for instance, is inclined to attribute a larger share of the differences to pre-natal factors than is Freeman. One of the authors is inclined to treat the unseparated identical twins as a control group and to regard only the excess differences of the separated twins as attributable to their post-natal environment.

When all nineteen cases are included, the separate and joint influences of educational and social differences account for seventy-two per cent of the variance in IQ difference; if four extreme cases are omitted, they only account for twenty per cent of the IQ variance. "The analysis indicates that the rôle of heredity and environment in producing twin differences is a function of the type of environment. Thus for twins reared together, most of the difference between members of a pair may be due to the nature factor; whereas for twins reared under strikingly different environments, the nurture factors will have a relatively greater influence. It is apparent from several of the comparisons made that the relative effect of hereditary and environmental differences is also a function of the type of trait. Any fixed ratio of these two factors for all traits and conditions is thus impossible. From the viewpoint of the educator it is important to note that extreme differences is educational and social environments are accompanied by significant changes in intelligence and educational achievement as measured by our tests" (p. 349). This is frightfully disappointing. Has the mountain labored for ten years to produce such a mouse? However, to discover that the problem is incapable of a general solution, since there is not one prob-

lem but a multiplicity of minor problems, is perhaps worth while. In this reviewer's opinion, however, we should not rest at this stage. Somebody with greater skill and acumen may be able to take the data and work it out to more clear-cut conclusions. We cannot remain satisfied with Jennings' dictum that what heredity can do, environment can also do. It runs contrary to common sense and to common observation, and anything that runs contrary to these is always under suspicion. Is the work of Morgan and his school as insecure in its foundations as this work would make it out to be?

PETER SANDIFORD.

University of Toronto.

GERTRUDE HILDRETH. *Learning the Three R's.* Minneapolis: Educational Publishers, 1936, pp. IX + 824.

This text, designed for the educational psychologist and the student of educational methods in the elementary school, is confined to a discussion of reading, writing, spelling and arithmetic. In successive sections of the book are considered, respectively: (1) Description of the newer program in the four skills; (2) the problem of school failures; (3) diagnosis in the four skills; and (4) suggestions for improving techniques of instruction.

The author is to be especially commended for two major points of emphasis: (1) The genetic approach, which is based upon the view that "learning is most successful when the task is adapted to the capacities of the child or suited to his maturation level." The recent increased interest shown in readiness for reading illustrates the progress of this point of view among teachers and school administrators. (2) Flexibility of instructional techniques is favored rather than any single method, such as textbook teaching and the like. Thus a more casual, informal, and natural introduction to learning results.

The section on diagnostic and remedial methods is by far the best part of the book. The evaluation of techniques is critical and the suggestions on programs, materials, sources of information and interpretation are adequate.

Many readers of this book, especially the students, will be dissatisfied with the organization of the first two-thirds of the material. The author often adopts an approximation of the outline method. She frequently presents in itemized order practically all that has been said or found out about a subject. An example is the listing of thirty-three

suggestions for wholesome personality adjustment in the school child. Most readers would get more from a discourse based upon evaluation and correlated discussion of the items. The subject-matter in the field, however, is very completely covered.

The typography of the book is excellent with one exception; the highly glazed paper produces an uncomfortable glare in most artificial light.

For perhaps the first time one finds in a textbook a satisfactory evaluation of the rôle of eye movements in reading. "Eye movements are resultants of perceptual habits which are in turn dependent on comprehension patterns." There is, however, an unfortunate departure from this point of view when it is stated that reversals of letter sequence in perceiving certain words are due to faulty eye movements, and suggestions for training the "eye sweep" to remedy the situation are given.

This book will be rated high by most educators and will undoubtedly find much use. It is difficult to imagine a more complete covering of this particular subject-matter.

MILES A. TINKER.

University of Minnesota.

The Psychology of the Common Branches

By WILLIAM H. PYLE

"Doctor Pyle has produced a very useful, teachable, and productive contribution to the psychology of teaching. In this book the principles essential to the effective teaching of reading, spelling, handwriting, and arithmetic are set forth . . . The book is not only one that can be used as a textbook but it is an exceptionally good reference work."—*Journal of Educational Research.*

"The book is written in clear, non-technical style and should be very valuable for the student in training or for the teacher who wishes to make a quick survey of the scientific principles involved in teaching the four skill subjects."—*Bulletin, New York Society for Experimental Study of Education.*

$2.00 *plus* 8¢ *postage*

WARWICK AND YORK
Publishers Baltimore, Md.

Warning!

When You Say Webster's Dictionary
You MEAN the MERRIAM-Webster

Accept Nothing Less Than the SUPREME Authority"

WEBSTER'S NEW INTERNATIONAL DICTIONARY Second Edition

LOOK FOR THIS CIRCULAR TRADE-MARK AND THE "MW" MONOGRAM ON THE COVER.

THROUGH four generations Webster's Dictionary has earned and maintained first place among reference books. When you are confronted with any question of *fact*, whenever you need to *know*, you think of *Webster* as the utmost in authority, the judge of judges, the court of last appeal. *But that confidence is warranted only if you use the Merriam-Webster.*

Look for the Merriam-Webster trade-mark when you buy a dictionary. The *New* Second Edition of Webster's New International (unabridged) is 20 years newer, contains 122,000 more entries than any similar dictionary. Provides information in all branches of knowledge. The authority in courts, colleges, newspapers. Prepared by 207 experts. 600,000 entries; 3,350 pages. 12,000 terms illustrated. Ask your bookdealer, or write for free booklet.

G. & C. Merriam Co., Dept. 11, Springfield, Mass.

A DETERMINATION OF GENERALIZATIONS BASIC TO THE SOCIAL STUDIES CURRICULUM

By NEAL BILLINGS

PRESENT conventional courses in history, geography, and civics fail to achieve as adequately as possible the great aims of the social studies. Granting that existing courses acquaint pupils with many of the main facts of life, still not all of the important facts are presented nor are the vital inter-relationships indicated.

In this study Professor Billings points the way toward a more effective curriculum by setting forth the basic generalizations needed by those who are making curricula in the social studies in order that prospective American citizens may be taught to think intelligently about the conditions and problems of contemporary life.

$3.00, postpaid $3.12

WARWICK AND YORK, INC.

Publishers BALTIMORE, MD.

educational psychology in *the Spring Term*

by **W. H. Pyle**

THE PSYCHOLOGY OF LEARNING. Revised and Enlarged Edition. ***$2.20; postpaid $2.32.***

"Professor Pyle has performed a prodigious task wonderfully well. The ground has been well covered and the conclusions impartially stated. A whole literature of learning has suddenly been made accessible to the teacher." Peter Sandiford in *The School*, (Toronto).

THE PSYCHOLOGY OF THE COMMON BRANCHES. ***$2.00; postpaid $2.10.***

"The book is written in clear, non-technical style and should be very valuable for the student in training or for the teacher who wishes to make a quick survey of the scientific principles involved in teaching the four skill subjects." *Bulletin, New York Society for Experimental Study of Education.*

by **E. K. Strong, Jr.**

INTRODUCTORY PSYCHOLOGY FOR TEACHERS. Revised and Enlarged Edition. ***$2.20; postpaid $2.32.***

Instead of beginning with the most uninteresting phases of psychology and those least familiar to students, this course takes up concrete experiences of everyday life, relates them to the problems of learning and individual differences, and so develops these two topics.

BRIEF INTRODUCTORY PSYCHOLOGY FOR TEACHERS. ***$1.60; postpaid $1.70.***

"A large field of usefulness is predicted for this book and its influence is bound in time to extend to the whole field of textbooks in education." *The Virginia Teacher.* This book contains Part One of Strong's *Introductory Psychology for Teachers* and is designed for use in a short course.

by **A. S. Edwards**

THE FUNDAMENTAL PRINCIPLES OF LEARNING AND STUDY. Revised Edition. ***$1.80; postpaid $1.90.***

"The form of presentation is clear and pointed, so that the book might be used as a text in a short course, or it might very well serve somewhat as a guide to study and discussion in a long course."—*American Journal of Psychology.*

by **Howard Taylor**

INTRODUCTION TO EDUCATIONAL PSYCHOLOGY. ***$1.60; postpaid $1.68.***

The book has a flexibility of organization that makes it easily adaptable to the particular needs and interests of both teacher and class. It may be used as a basic text or to supplement some other introductory text.

WARWICK and YORK

Publishers **BALTIMORE**

THE MAPLE PRESS COMPANY, YORK, PA.

C2399
INCHES
METRIC